The Primacy of Phenomenology Over Cognitivism
Towards a Critique of the Computational Theory of Mind

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Graduiertenschule für die Geisteswissenschaften der
Julius-Maximilians-Universität Würzburg

vorgelegt von
Jethro Masís

aus San José, Costa Rica

Tag der mündlichen Prüfung: 02. Februar 2015

Vorsitz des Promotionsprüfungsverfahrens: Prof. Dr. Brigitte Burrichter
Gutachter und Erstbetreuer im Promotionskomitee: Prof. Dr. Karl-Heinz Lembeck
Gutachter und Zweitbetreuer im Promotionskomitee: Prof. Dr. Karl Mertens
Zweitbetreuer im Promotionskomitee: Prof. Dr. Franz-Peter Burkard
Protokollanten: Dr. Robert Ziegler

Meinem Sohn Andreas (Würzburg den 24.11.2011) zugeeignet.

"Ils accumulaient des données, lourdes et répétitives, dans le seul but d'en tirer des applications industrielles immédiates, sans jamais prendre conscience que le socle conceptuel de leur démarche était miné."

Michel Houellebecq
Les particules élémentaires

Contents

ACKNOWLEDGEMENTS

Work leading to the completion of this work has taken me from Costa Rica to Germany and understandably such journey would not have been possible without the financial and logistic support from various institutions and the encouragement of a great number of individuals (the majority of which I won't be able to list right now).

This is a (nonsubstantially) revised version of the doctoral dissertation that was submitted and defended at the Julius-Maximilians-Universität Würzburg and whose primary advisor was Prof. Dr. Karl-Heinz Lembeck (Chair of Theoretical Philosophy). Special thanks go to professor Lembeck, for his constant support during my stay in Würzburg, his excellent mentorship, and above all for his always friendly attitude and his hospitality. To him I am also indebted theoretically, since I have profited from his own investigations concerning the role of philosophy over against the hegemonic canon of the sciences. I also thank Prof. Dr. Karl Mertens (Chair of Practical Philosophy) and Prof. Dr. Franz-Peter Burkard (Program of Philosophy of Religion), for accepting to read previous versions of this investigation and being part of the doctoral committee.

My stay (and my family's) in Germany has been possible thanks to a scholarship for doctoral studies awarded by the German Academic Exchange Service (DAAD). My thanks go to Veronica Metje from the DAAD, who was always supportive and who provided me with all the necessary information about being a scholarship holder. I was also partly funded by the Universidad de Costa Rica through the auspices of the Oficina de Asuntos Internacionales y Cooperación Externa. From that office, I thank Silki Pérez, who was in charge of my institutional link to the Universidad de Costa Rica. During these years, several offices at the Universität Würzburg have been unfailingly hospitable. I would like to thank Dr. Thomas Schmid, Managing Director of the Graduate School of the Humanities at Würzburg, who kept me informed of all the details concerning the doctoral program at the Graduate School. I also thank Sybille Deuerling, secretary of the Chair of Theoretical Philosophy, and Ursula Shahmary from the International Office.

I thank all the friends I met during this four-year period in Germany. Special thanks go to Karl Vogt in particular for his warm friendship and his closeness to my family. Special thanks must also be

given to Paola Ravasio, who more than a friend, has been part of my own family. I thank her also for reading several versions of the manuscript and suggesting several stylistic changes.

San José, July 2016

| 1 |

PROLEGOMENA TO COGNITIVISM

The Panorama

The story of the confrontation between phenomenology and cognitivist philosophy which is here presented might seem as if told from the perspective of an objective historian, which often makes it hard to tell when the narrative is depiction of elements of factual history or rather a narrative which is partially oriented. But there is not such narrative of facts in the history of theoretical ideas, so let there be no mistake about it: my story is so arranged in order to accomplish the task of examining fundamental ideas of current cognitive science and philosophy of mind from the point of view of phenomenological philosophy. Thus, my perspective is not neutral nor is it the use I make of the debates resorted to in the configuration of this investigation.

Above all, this is a story of philosophical problems staging discussions whose guiding thread—the confrontation between phenomenology and cognitivism—does not necessarily exist already in the literature and when it does, the perspective adopted is rather foreign or not precisely interchangeable with the concerns motivating this investigation. Therefore, my narrative is so crafted in order to show how phenomenology has been received within cognitive science but it also has a normative purpose: to imply how it should have been received and why that matters. This means that while my perspective might seem hidden, the whole narrative is framed along lines which can be said to establish the delimitation of philosophy (phenomenology in this case) from the sciences, particularly the neuro-cognitive and informational ones. But also this story so concocted attempts to show how the paradigm embracing all these sciences, that is, cognitivism, is in itself not devoid of philosophy but rather constitutes a philosophical bent, although this is not always recognized as such. In this sense, my active role as a researcher consists in imagining relationships between theoretical discourses which are not necessarily given, so I see my work as the task of uncovering surreptitious similarities and repulsions between several approaches in order to make visible what was not noted, or not good enough. My

purpose is to unravel the tangled skein which constitutes the contemporary discourse on how to assign its rightful place to mind in nature and how phenomenology tackles this problem, or rather debunks it.

In the pages that follow, I argue that philosophy is irreducible to the aforementioned sciences and that something fundamental would be completely gone forever if it were possible to annul philosophical questioning, that is, if it were conceivable to reduce it to some more basic science or to translate it into cognitive jargon without remainder. I agree with Deleuze that "philosophy is at its most positive as a critique, as an enterprise of demystification" (2006, p. 106). Therefore, the Deleuzian question remains: "is there any discipline apart from philosophy that sets out to criticise all mystifications, whatever their source and aim, to expose all the fictions without which reactive forces would not prevail?" (*idem*). Since the scope of my investigation is framed along the dreams of cognitivist philosophy of mechanizing the mind and naturalizing phenomenology—a program already put into operation by Quine's demand to naturalize epistemology (see 1969, p. 69 ff)—the reversal of the high expectations of the so-called 'cognitive revolution' (see Gardner 1987) concerning its purpose of elucidating fundamental philosophical questions should be sought. This is why my investigation can be construed as an attempt to demystify the cognitive fictions pervading much of contemporary philosophy of mind: for instance, that cognition can take place without a body; that consciousness can be naturalized and that it is just the final emergent ingredient in the recipe of cognition or, worse, an illusion generated by a representational biological machine; that the brain thinks, feels, and makes decisions; that the world is a collection of discrete facts which can be abstracted away in a computer program; and that being-in-the-world is, indeed, such computer program. Accordingly, I shall argue that cognitive science is not philosophically neutral and that the most traditional branch of cognitivist philosophy supposes many of the tenets that were entertained by modern philosophers.

As is widely known, however, this sort of critique has already been made (see Dreyfus 1992) and for much the same reason, the novelty of this investigation must reside somewhere else. My idea is, then, to offer also a revisionary interpretation of those early phenomenological claims against the precious toy of cognitive science: artificial intelligence (AI henceforth). And then inquire (to later answer in the negative) whether Dreyfus's critique of artificial reason is phenomenological enough. Furthermore, more recent approaches attempting an integration of phenomenology into the cognitive

scientific enterprise shall be discussed in depth. I seek to show that phenomenology should guard itself from false friends that might end up confusing its 'matter of thinking' or *Sache des Denkens*. Doubt is thus cast on the relegation of phenomenology sheerly to the province of the first-person perspective. And the legerdemain as to the program of naturalizing phenomenology will be exhibited as rendering itself incoherent in the process.

In order to tell the story the way I have planned to since I first imagined the structure of this investigation, I have had to draw from discussions having their origin in the philosophy of mind, the philosophy of psychology, cognitive science, and phenomenology. However, my purpose is more to extrapolate from these discussions than to expound them. Therefore, this is not an exegetical work in which the attempt is made to arrive at criteria by means of textual evidence. So I do not strive to solve philosophical debates, although I shall exhibit some of them and will be explicitly willing to take sides. Drawing from discussions stemming from all the aforementioned theoretical discourses might require a very selective procedure in identifying what needs attention and what not. The procedure is admittedly selective given not only the copious literature touching upon these topics but also the enormous theoretical scope of the fields. In order to leap over this obstacle or otherwise getting around how to manage all this information, the aspects of this massive theoretical panorama that matter need to be decided from the outset. And that I have done as follows.

The first section, *Dreyfus and Cognitive Science*, revises the philosophical origins of cognitive science precisely from the perspective of Dreyfus's major work *What Computers Still Can't Do* (1992; originally 1972). His later work can be seen as a constant dialogue with his first book, so in a way the main traces of his whole work are discussed in depth and I put them in context with former and current debates. As such, the discussion is organized around Dreyfus's arguments and the counterarguments that were fired at him when he first attempted to assess philosophically an enterprise that thought of itself as merely technical in nature and vocation. The consequences of Dreyfus's critique as the first introductor of phenomenology into cognitive science discussions will be unfolded up to the point where analytic philosophers of high stature (Dennett, Searle, McDowell, Haugeland, etc.) cared to intervene. This will permit a thorough presentation of Dreyfus's main philosophical ideas

as well as the theoretical hardships and shortcomings in which he fell, due in part to his lack of familiarity with the rich phenomenological tradition and his rather selective appropriation of it.

Dreyfus's role in the whole history of the philosophical debates within cognitive science and the philosophy of mind is not to be underestimated and I explicitly intend to remark its importance. Evidence of Dreyfus's enduring repercussion can be found in how his disciples or otherwise researchers partially influenced by his work are scattered around the world. It is not a mere irony, but a confirmation of his shrewdness, that some theoretical problems of the cognitive enterprise that were soon enough (in the distant 1970s) highlighted by Dreyfus, actually happened to turn real decades ahead. Nevertheless, he is not only the introductor of phenomenology in Anglo-American philosophy but also a translator—however much he may wish he were not—of phenomenological concepts into pragmatist and cognitive ones. In addition, he has also been instrumental in shaping a simplified version of the history of philosophy beginning in Plato and traversing with haste through Modernity where Descartes founded the modern conception of mind. If one merges the theory of objectivity founded by Plato with Descartes's philosophy of mind, one gets as a result almost every philosopher of the twentieth century (including Husserl) because, according to Dreyfus, it was not until Heidegger's breakthrough with *Sein und Zeit* that a genuine critique of metaphysics challenged for the first time the traditional tenets of philosophy. This simplification of the history of philosophy will be corrected throughout my investigation, in particular by exhibiting the close connection between Heidegger's philosophy and that of the founder of phenomenology, Husserl. That master and disciple practiced phenomenology differently does not in the least affect the indisputable fact that both were practicing phenomenology.[1] Internal phenomenological debates (for instance, Heidegger's well-publicized *Auseinandersetzung* with Husserl) will either be alluded *en passant* or glossed over altogether. And this for the sake of a greater good: namely, providing a united phenomenological front against the claims of cognitivism.

The second section, *Phenomenology and Cognitive Science*, evaluates in depth the uses and misuses of phenomenology in cognitive science. The analysis will be carried out beginning with the contemporary

[1] Of course, Heidegger was not Husserl's direct disciple. However, "aufgrund des tiefen Einflusses, den das Husserlsche Denken auf Heidegger in der Zeitspanne fast eines Jahrzehnts ausgeübt hat, kann man trotzdem mit einem gewissen Recht von einer Lehre-Schüler-Beziehung sprechen" (Cristin 2012, p. 44).

discussion concerning the philosophy of consciousness, since this will pave the way for the procurement of an understanding of the appropriation of Heideggerian philosophy in cognitive science, which is very ample and variegated. Indeed, Heidegger's explicit reception in cognitive science circles includes the design of behavior-based robots; applications in human-computer interaction, entrepreneurship, and the social practices surrounding computing; the attempt to program *Zuhandenheit* based on a critical technical practice which takes responsibility with the metaphors pervading cognitive science; and a neo-Heidegerian turn which admittedly calls itself 'Heideggerian' while attempting to get rid of the transcendental and antinaturalist strands so characteristic of phenomenology. The root of the matter is that Heidegger's *Daseinsanalytik* has been redefined and translated into cognitive terms. Dasein, for example, has been turned into 'human agency' and Heidegger's analysis of the worldliness of the world (*Weltlichkeit der Welt*) is thought to be apt for developing computational theories of human interaction as much as for demolishing the 'Cartesianness' of much work in cognitive science. The cognitive Heideggerians expect that by incorporating explicit Heideggerian insights into cognitive work, cognition can secure the right ontology and avoid the theoretical stalemates and impasses to which cognitive science has frequently fallen prey. My purpose, however, is to put this aspiration to the test.

The last section, *Conclusion*, amalgamates the topics of the whole investigation and brings forth the last resort of a phenomenological reception in cognitive science: that of the project of a phenomenological philosophy of mind which finds it appropriate to import heavily phenomenological insights to bear on technical discussions in the philosophy of mind. Here a balance is sought between empirically-informed inquiry and phenomenological philosophy, both of which are supposed to mutually constrain each other. That I shall find not to be a minor quibble but rather an intricate business which deserves a critical take.

I admit that, given the selective character of the discussions which form part of this investigation, some other debates and some other works could have been perfectly dealt with. I constantly decided on the run not to refer with significant extent to other topics and debates that could also have had some phenomenological import. However, it is a tragic circumstance of this kind of research that part of it is to delimitate its boundaries. I hope, nonetheless, that the confines that were

chosen are wide enough in such a way that they do not inhibit the full scope which is needed for achieving the purposes of this investigation.

Cognitivism

In view of the wide ramifications of the themes which are to be touched upon in due course, a few words regarding the tradition of thought which shall become the object of staunch criticism are needed.

I already mentioned that cognitive science—much to the disappointment and retorts of many of its practitioners who do not (and cannot) see themselves as being guided by philosophical ideas—is not philosophically neutral. Cognitive scientists might not be encumbered by such theoretical bottlenecks as the ones found in philosophy, in particular because they are technical people working in laboratories and environments suited for the purposes of achieving empirical results. Consequently, such scientists think of themselves as being in the opposite end with regard to 'armchair philosophizing.' But that does not preclude the fact that cognitive science has had philosophical aspirations from the start. And this is due to cognitivism: the philosophy behind much research in cognitive science. Cognitivism is rooted in the computer as a metaphor for thinking and, as Dupuy has argued, it resulted "from an alliance between cognitive science and the philosophical psychology that is known as philosophy of mind, currently a very active branch of analytic philosophy" (2009, p. 12). I also submit Dupuy's thesis that "what gives coherence to the many research programs that go under the name of cognitive science today is the philosophical work being done in connection with them" (2009, p. 90). For much the same reason, it is cognitivism or 'the new mental philosophy' (see Descombes 2001, p. 66 ff) what entails the classification of different sciences under the rubric of being 'cognitive.' The latter point explains why the cognitive revolution has caused controversy and excitement in about equal measure, since it could be construed either as the key for breaking the Gordian knot of the most elusive problems of science and philosophy or simply as a reductionist stratagem. Descombes discloses the cognitivization of the sciences as follows:

> If the sciences called 'cognitive' were simply those whose objects are intellectual ones like language or culture, then the word 'cognition' would be understood in the ordinary, precognitivist way, an understanding to which notions like the treatment of information and artificial intelligence are utterly alien. Yet when the cognitive theorists include already-existing disciplines

> among the cognitive sciences, they do so in order to propose that these disciplines redefine themselves in the light of a new conception of mind. Thus construed, cognitivism is the program that seeks to change the sciences of the mind into cognitive sciences. For example, if linguistics counts among the cognitive sciences, it is not because it studies languages, which are intellectual systems. Rather, linguistics counts as a cognitive science because it could, we are told, be redefined or reconstructed as psycholinguistics, the study of the linguistic capabilities of a mental system endowed with the 'organs' necessary for the understanding and production of sentences. (2001, pp. 66-67)

Indeed, 'cognitive' is nowadays an epithet whose usage is omnipresent. And this not only regarding the usual sciences included originally in the cognitive turn, such as neuroscience, psychology, anthropology, linguistics, and computing. Now ecology (see Dukas & Ratcliffe 2009), religion and theology (see Boyer 2008; Barrett 2011), theater, performance, and acting (see McConochie 2008; Lutterbie 2011), poetics and literature (see Stockwell 2005; Jaén & Simon 2012), archaeology (see Abramiuk 2012), and even medicine (see Shortliffe & Cimino 2014) are said to be cognitive or they are regarded as being at least approachable from a cognitive perspective. It should not surprise anyone that even the social sciences (see Zerubavel 1999; Levy 2013; Bracher 2013) and the arts (see Massey 2009), which could be thought to be more prepared for resisting reductionisms of any sort, are now genuflecting to the cognitive. So cognitive sociology, cognitive politics or cognitive history are not anymore bizarre expressions but more and more prevalent ones.

With regard to philosophy, I hold that it is suitable to extend to this whole process of cognitivization Proust's (1987) idea that there is an 'implicit philosophy' of AI. So there is, I argue, an implicit philosophy of cognitive science, cognitivism, which also has more than just empirical aspirations: it is a transcendental hypothesis about reality. Indeed, as Proust has convincingly argued, AI as a research program embodied from the start a philosophy of transcendental type according to which the formal conditions of cognition common to all systems that are cognitive (whether humans, animals, or machines) had to be sought. Thus, the purpose was to uncover the a priori and necessary conditions constituting any cognitive system. As Dupuy argues, "the transcendental subject was replaced by the 'physical symbol system' and the universality of the synthetic a priori by the universality of the Turing machine" (2009, p. 93). Whatever the differences, adds Dupuy, "the distinction between psychology and the critique of knowledge, between contingent laws of cognition and necessary rules, was carefully made" (*idem*). One is, therefore, not to be fooled by the typical discourse on account of cognitive scientists that they are solely faithful to the methods of the empirical sciences, and nothing more.

At any rate, the relationship between cognitive science and philosophy is problematic and not easily admitted. Following Franchi (2006), one is driven to the conclusion that something somewhat paradoxical happened between cognitive science and philosophy: AI arose precisely in a self-conscious opposition to the methods of classical philosophy and turned toward the sciences ir order to find the right tools of analysis and a methodological cannon. But it ended accepting the need for an all-encompassing theory of human nature: "it saw itself as a 'new philosophy' and indeed as an anti-philosophy that aimed at recovering the goals and scope of the millennia-old attempts toward an exhaustive account of man and his place in the cosmos, while replacing armchair speculations with a radically new kind of empirical approach." (Franchi 2006, p. 27). Cognitive science as a research program can then be described in the following perplexing terms: cognitive science is both philosophical and antiphilosophical since it has always wanted to become the true philosophy, although by turning its back on philosophical speculation.

The present investigation intends to deal in depth with these claims and to draw out some of the morals deriving from them.

| 2 |

THE PHILOSOPHICAL ORIGINS OF COGNITIVE SCIENCE

AI and Philosophy

History spanning the different stages of AI has been narrated several times and in many forms, whether historical, journalistic, technical reporting, etc.[2] And one of these approaches is, of course, the philosophical.[3] Although it might appear as if taking a philosophical stance toward problems dealing with quite complex scientific and technical matters is not the most interesting to the field itself —for the field strives exclusively for technical achievements— the history of AI is full of rather extravagant predictions concerning the collection of capacities that we usually ascribe only to humans being chipped away in the near future by the application of a research program based on the functionality of computers. Indeed, computational and information-theoretic approaches in philosophy have grown out increasingly from the computer revolution, and some think that research inspired in these approaches "revitalises old philosophical questions, poses new problems, [and] contributes to reconceptualising our world views…" (Floridi 2002, p. 117).

However, the idea must be stood on its head, for the appreciation showed by philosophers to these new technical fields (that they regard as having some purchase on philosophy itself) is not always reflected in a corresponding interest on behalf of scientists for philosophy. Perhaps it is no surprise that technical people might be inappreciative of armchair (philosophical) discussions —discussions that, without the proper technological implementation, might seem futile. Certainly, nobody expects that a substantive intervention in the field, a theoretical breakthrough, could be in the offing on account of philosophical speculation. And his might be the reason why philosophical 'in-principle arguments' are uninteresting for technological purposes. This also explains the tendency in technical

[2] See McCorduck 1981; Gardner 1987; Dreyfus 1992 and 2012; Crevier 1993; Boden 1995 and 2006; Franci & Güzeldere eds. 2005; Lungarella et al. 2007; Steels 2007; Schmidhuber 2007; Husbands, Holland & Wheeler eds. 2008; Nilsson 2010.

[3] See Haugeland 1989; Boden 1990; Fetzer 1990; Cummins & Pollock eds. 1991; Copeland 1993; Dyson 1997; Haugeland ed. 1997; Gams, Paprzycki & Wu eds. 1997; Carter 2007; Riskin ed. 2007; Nath 2009.

parlance to construe philosophy and speculation as feeble and precarious, that is, as inapt and ineffective with regard to attaining technical results. A thorough philosophical position might ensue as a result of technical implementation but it must never stand in the way of technical research. So this is a narrative according to which, from the very outset, the goals and assumptions of AI strains the relationship between philosophy and its research program. However, as should be explained in detail in due course, philosophy strives to achieve wholly different comprehensive goals and it does not share the same topics and theoretical stance characteristic of the sciences (artificial or natural). As a matter of fact, the relationship between philosophy and the sciences should not simply be mentioned *en passant*, for it is part and parcel of a philosophical approach to the cognitive turn. So while there is a technical practice that suspects of philosophical speculation, there reigns also the idea of a quasi Hegelian *Aufhebung* of philosophy resting in the hands of technological practitioners. Among technological scientists, indeed, claims according to which old metaphysical conundrums could be solved by means of the right technological implementation are often entertained and taken for granted, as though the task of science were to overcome old philosophical dreams by way of making them come true.

Philip Agre has summed up a common claim from the AI community in which its highly technical work is judged—or rather misjudged, depending on how one looks at it—from the point of view of philosophical inquiry: "technical people reject the applicability of philosophical analysis to their activities, arguing that practical demonstration forms a necessary and sufficient criterion of success for their work" (1997, p. 23). Therefore, the question could be raised as to what extent a philosophical reflection upon technical matters constitutes an intervention of exogenous nature in the field. If philosophers charge AI researchers with practicing philosophy and not merely technical work, it could be rebutted that this constitutes the suspicious undertaking of condemning what is not the case, or, even worse, of attributing mysterious goals unbeknownst to the researchers in question. AI researchers retort that such philosophical claims miss the point of their whole effort, which cannot be challenged —let alone refuted—unless on technical ground.

Indisputably, concentration on technical success explains the fact that AI practitioners do not recognize themselves as carrying out the philosophical ideas that critical assessments of AI coming from philosophers ascribe to them (Agre 1997, pp. 21-22). Therefore, they do not think of themselves,

for example, as Cartesian, never mind as dualist, or in any case as realizing by technical means philosophical ideas of any kind. Conversely, their concern focuses on building machines and robots, creating experts systems, and writing programs that can be implemented in some kind of hardware, or that can have some industrial applicability, and hence they think of themselves as less committed to philosophical ideas than are philosophers. According to this narrative, computer scientists working on AI would without hesitation change their approach to any given problem if it suits the technical results they are trying to attain, and purported philosophical tenets would thus be swept aside if empirical results would suggest they must. There would be, then, no commitments of any kind being dogmatically entertained: explanatory success through technological implementation is the only guiding thread of investigation. Accusations as to whether a philosophical tenet is being nonetheless held could be explained as incidental and not as substantive to the technical work being carried out. The least that could be said is that AI researchers do not willfully embrace ideas coming from speculative philosophy circles, and, even if it seems like it, it is more a matter of coincidence than of actual philosophical commitment. The attempt to make this clear explains the passing (although startling) statements by roboticist Rodney Brooks of MIT that his approach to robotics has not been influenced directly by German philosophy: "in some circles much credence is given to Heidegger as one who understood the dynamics of existence. Our approach has certain similarities to work inspired by this German philosopher… but our work was not so inspired. It is based purely on engineering considerations" (Brooks 1999a, p. 97).

However, such a way of drawing the frontiers between the practical/technical and the philosophical/speculative is highly problematic. On the one hand, AI is a quite distinctive interdisciplinary field which from the start nourished its imagination with ideas coming from different disciplines such as engineering and cybernetics, biology, experimental psychology, communication and game theory, mathematics, logic, and, most certainly, philosophy. As has been handsomely underscored by Varela, Thompson & Rosch in their classic work (1991, p. 38), the use of mathematical logic to understand the nervous system, the invention of information-processing machines, the establishment of a metadiscipline such as systems theory, along with the conception of information as a statistical theory and the first examples of self-organizing systems; all these elements of the cybernetics phase of

cognitive science helped to give shape to this promising science, which had AI as its heart. In addition, history of AI imagination can be traced all the way back to ancient literature and myth, and nowadays its transhumanist inspiration has been replicated in film and cyberpunk literature. No wonder, as one definition would have it, AI could be defined as "the science of how to get machines to do the things they do in the movies" (Astro Teller [as quoted by Kurzweil 1999, p. 66]). As Boden (1988) has keenly argued, psychological research based on computer-modeling is well rooted in culture:

> Working models of living creatures are not new. They have existed as chic toys for many centuries: in the palaces of ancient Alexandria, the courts of medieval Islam, and the estates of the eighteenth-century European aristocracy. For over a hundred years they have cavorted in the pages of fiction, focussing cultural fears and fantasies of various kinds. (p. 1)

Precisely because AI is not solely a robotic or engineering enterprise, but "also about understanding the nature of intelligent thought and action using computers as experimental devices" (Buchanan 2005, p. 54), it is perhaps understandable that its earliest practitioners conceived of their theoretical endeavors as supplanting those of sheer speculative philosophy. This might suggest that, contrary to what technical purists want us to believe, AI and philosophy have been closely related from the very outset of its journey toward replicating human intelligence. After all, "philosophy is the hidden framework in which all AI is done. Most work in AI takes implicit philosophical positions without knowing it" (Chapman et al., 1988).[4] As Dennett asserts accordingly: "AI is, in large measure, philosophy. It is often directly concerned with instantly recognizable philosophical questions" (1988, p. 283).[5] And so we are told by perhaps the first harsh critic of the project, Hubert Dreyfus (sometimes referred to as 'the black knight of AI', [Dreyfus 1992, p. xvii] or 'the ideological foe of AI', [Dennett 1984, p. 131]), that when he was teaching at MIT in the 1960s, students from the AI laboratory came to his Heidegger course claiming as follows: "You philosophers have been reflecting in your armchairs for over 2000

[4] This reference, quoted as Chapman et al. 1988, was actually edited by Chapman and appears to have been written by "a whole bunch of current, former, and honorary MIT AI Lab graduate students." For details, see the URL provided in the reference section.

[5] "What is mind? What is meaning? What is reasoning and rationality? What are the necessary conditions for the recognition of objects in perception? How are decisions made and justified?" (Dennett 1988, p. 283). This idea that AI is philosophy is, of course, not universally agreed upon. Putnam, for example, thinks AI is simply a subbranch of a branch of engineering, namely, computer science, and deems it preposterous when AI researchers claim they are practicing epistemology (see 1988, p. 270).

years and you still don't understand intelligence. We in the AI Lab have taken over and are succeeding where you philosophers have failed" (2007a, p. 247).[6]

Indeed, if technological means are now available that can actually test what cognition is all about, that is, if there is a way to realize cognition technologically instead of just discussing about it, all those apparently unproductive epistemological and ontological disputes typical of philosophers could be deemed superfluous once and for all. According to the aforementioned technocratic assumption, philosophers have been arguing for centuries over the nature of the human mind and its alleged mysterious workings, and they have not reached so far an agreement. Now it is time to advance the realization of mere speculation through technological means, and computers are powerful tools which could lead to ruling out vane hypotheses in practice, not in thought. Philosophy appears, then, at the earliest stage of AI history although under the demand on behalf of field technicians to not get in the way. It was a matter of *realizing* philosophy, of making it come true; a task which in turn postulated, as it were, the tragic disappearance of philosophy itself on behalf of its realization. In this sense, replicating intelligence by technological means amounts to answering seminal questions on the nature of mind, consciousness, and on what supposedly constitutes human beings essentially: cognition. No wonder current parlance on the death of philosophy—the recent book by Stephen Hawking and Leonard Mlodinow, *The Grand Design*, is more than exemplary, although in the field of physics (2010)—toys with the idea that technical success would eventually quieten down much heated philosophical debates belonging to the past. How so? By way of achieving what philosophers have only dreamed of and discussed about.

One must concede, however, that this dismissal of philosophy was just a candid *desideratum*—as is, I think, the failed case advanced by Hawking and Mlodinow of overcoming philosophy through physics—and all this talk about the overthrowing of philosophy by engineering appears exaggerated, if not overtly chimerical. As a matter of fact, back in the 1960s Dreyfus witnessed first hand that far from finding in effect this purported rejection of armchair philosophy in the field, he encountered what he thought was a putting-into-practice of well-known philosophical theses, such as the following:

---

[6] The first clouds on the AI horizon, though, appeared early in the 1960s. "As early as 1961, an engineer named Mortimer Taube published the first anti-AI book under the title *Computers and Common Sense: The Myth of Thinking Machines*. Perhaps because its author died shortly after publication, the book raised few eyebrows" (Crevier 1993, p. 120).

> Hobbes's claim that reasoning was calculating, Descartes' mental representations, Leibniz's idea of a 'universal characteristic' (a set of primitives in which all knowledge could be expressed), Kant's claim that concepts were rules, Frege's formalization of such rules, and Wittgenstein's postulation of logical atoms in his Tractatus. In short, without realizing it, AI researchers were hard at work turning rationalist philosophy into a research program. (*idem*)

To a certain extent, this Dreyfusian discovery of masked philosophical assumptions and, in some way, science fiction dreams wanting to disguise themselves as unbiased and purely technical work, gives us a glimpse of what a philosophical stance on the history of AI pretends to tackle. Isn't it ironic that it is actually the scientists the ones pursuing a sort of empirical metaphysics by trying to understand intelligence apart from embodied situatedness? Or by attempting to reduce cognition, like Nobel Prize laureate Francis Crick (1994), to a vast assembly of nerve cells and their associated molecules? Many examples of this kind of reductionism flourish mightily. The computational one gives Haugeland a reason to affirm that scientists are "more often guilty of 'armchair philosophizing'... than are philosophers like Dreyfus, [because it is] typically the *opposition* to the necessary situatedness of intelligence that has been based on *a priori* presuppositions" (1996c, p. 120); a metaphysics that lurks unexamined in the unchampioned assumptions of cognitivism, which will be discussed in depth in due course.

So, from the point of view of a philosophical investigation, it has no relevance whatsoever if AI practitioners do not assume themselves as philosophers, but it is of utmost importance if certain philosophical assumptions are actually working—albeit under cover—precisely there where it is claimed that technical practice and speculation do not fit together. It might well be the case that the philosophical character of AI is elusive, in such a way that, as Wheeler has argued, "it is typically invisible to the external observer and even to the majority of working cognitive science scientists, for it is buried away in the commitments, concepts, and explanatory principles that constitute the deep assumptions of the field" (2005, p. 14). And again, a clue of what a philosophical stance on the history of AI amounts to can be found in the purpose of excavating and removing what has been buried away in this manner, that is, by revealing the very philosophical assumptions guiding technical work.

In spite of the astonishing naïvety that could be entertained by just posing the question as to whether one can practice science without granting certain background ideas, hardly anyone today would admit to working without theoretical assumptions of any kind. Yet, from the merely technical point of

view, it is perhaps just about whether or not the assumptions underlying technical work being done are the right ones, lead to successful results, and so on. The question could be raised: What else but technical success using computers as explanatory tools, could weed out false ideas about the nature of the mental? However, a purely internal understanding of AI—and this according to the well-known charges launched by Kuhn to normal science when he characterized it as mere puzzle-solving (1996, pp. 35-42)—could be easily objected, as was precisely the claim of normal science, perpetually accumulating a growing stock of puzzle-solutions, tending thus to avoid by all means the attempt at refutation. Furthermore, a sign of our times is the recognition of the all-encompassing effects of history and even prejudice—whose avoidance was once the clear gesture of the truly objective scientific endeavor—has been embraced as unavoidable, in fact, as a constitutive dimension of human life.[7]

However, in order to characterize more suitably the inevitable association between AI and philosophy, it suffices to stress that it all depends on which definition of AI one entertains. So what is AI? According to Margaret Boden (1990), there are at least four ways of coming to grips with a suitable definition. AI could be defined as

> · the study of how to build and/or program computers to enable them to do the sorts of things that minds can do.
>
> · making computers do things that would require intelligence if done by people.
>
> · the development of computers whose observable performance has features which in humans we would attribute to mental processes.
>
> · the science of intelligence in general, or more accurately… the intellectual core of cognitive science. (p. 1)

As the practical attempt to write programs displaying cognition, AI can be best understood as the core of cognitive science, that is, as the attempt to prove in practice those theoretical ideas coming from cognitive research, whose main goal "is to provide a systematic theory that can explain (and perhaps enable to replicate) both the general categories of intentionality and the diverse psychological capacities grounded in them" (*idem*). To this effect, AI is the engineering branch of cognitivism: AI is mind

[7] See Gadamer on this regard: "Darum sind die Vorurteile des einzelnen weit mehr als seine Urteile die geschichtliche Wirklichkeit seines Seins" (GA 1, p. 285).

design. According to Haugeland (1996a), 'mind design' is the endeavor to understand the mind in terms of its design or how it is built. But it is more than mere engineering implementation, for, in addition, it amounts to a sort of cognitive psychology. Consequently, AI is oriented

> more toward structure and mechanism than toward correlation or law, more toward the 'how' than the 'what', than is traditional empirical psychology. An 'experiment' in mind design is more often an effort to *build* something and make it work, than to observe or analyze what already exists. Thus, the field of artificial intelligence, the attempt to construct intelligent artifacts, systems with minds of their own, lies at the heart of mind design. Of course, natural intelligence, especially human intelligence, remains the final object of investigation, the phenomenon eventually to be understood. What is distinctive is not the goal but rather the means to it. Mind design is *psychology by reverse engineering*. (Haugeland 1996a, p. 1)

It should be very clear that early enough the association between AI and philosophy was evident. As Clark has argued, "the new sciences of the mind were to provide the long awaited vindication of the most potent dreams of naturalism and materialism" (1996, p. 1). However, the five decades which have already elapsed since AI's pristine inception as a research program speak of a variety of approaches ranging from classical AI, connectionism, situated and evolutionary robotics, artificial life, and the most recent embodied, embedded, extended, and enactive approaches.

The next step is to bring the origins of AI to the foreground, for a revision of them entails an understanding of the first difficulties and theoretical shortcomings that came about in the earliest phases of its research agenda.

The Origins of AI as a Research Program

Some bold statements advanced by Hubert Dreyfus in his seminal work *What Computers Still Can't Do* (1992) should suffice in order to stress what really is theoretically at stake with the ambitious technological research program embodied in AI:

> If we are on the threshold of creating artificial intelligence we are about to see the triumph of a very special conception of reason. Indeed, if reason can be programmed into a computer, this will confirm an understanding of man as an object, which Western thinkers have been groping toward for two thousand years but which they only now have the tools to express and implement. The incarnation of this intuition will drastically change our understanding of ourselves. If, on the other hand, artificial intelligence should turn out to be impossible, then we will have to distinguish human from artificial reason, and this too will radically change our view of ourselves. Thus the moment has come either to face the truth of the tradition's deepest intuition or to

> abandon the mechanical account of man's nature which has been gradually developing over the past two thousand years. (pp. 78-79)[8]

These are no doubt radical assertions, inasmuch as they seem to imply that AI puts us at a conclusive stage in which a decision is to be made as to whether the deepest intuition of Western philosophy, that is, objectivity (and eventually all its ramifications and potential in every discipline of knowledge), is essential. This is no trivial matter, for it is nothing less than the conception of what the human mind is what is at stake here. Moreover, facing this deepest truth, that everything can be objectified (even our consciousness), amounts to assessing to a greater extent the scope of the philosophical and scientific tradition of the West. But the danger that lurks beneath the promises of the mechanization of mind is that, if Dreyfus is right, a flagrant failure in AI research amounts to a very significant failure—and not only for engineering and technological reasons.

The reason why Dreyfus's work has been critically received to the point of mockery and has been, more often than not, plainly ignored, accounts for the fact that he touched some fundamental nerve at the explanatory level of AI. And he goes certainly very far in implying that the 'failure' he discovered at the heart of AI pertains not only to technology, but also and foremost to scientific theory as an enterprise of the West. The virulent reactions to his book are thus understandable since Dreyfus launched a series of attacks when the project was most promising: in its earliest phases and at the very heyday of the rise of computer machinery.[9] He launched a bomb, so to say, on the playground of naturalistic and empirical-oriented philosophers, whose target was their most precious idea: that of confirming by technological means the adequacy of a naturalistic framework; a desire that has cavorted in the minds of philosophers and scientists increasingly since the 19th century. Dreyfus warning is then twofold: the success in programming intelligence will be of great explanatory value with regard to understanding the human mind. But if the research program fails, precisely because of ill-suited

---

[8] The revised edition by The MIT Press, *What Computers Still Can't Do*—preceded by the original (1972) and second (1979) editions—adds 'still' to the original title, *What Computers Can't Do*. By this is meant, of course, that the fundamental results of the book are *still* valid, even forty years later.

[9] By 1972, when Dreyfus originally published his most famous work, the unimaginative use of computer systems characteristic of the 1950s and 1960s, with its valve-tubes and punch-cards, was already outdated. The microprocessor, the first 'computer on a chip' (as announced by INTEL in the early 1970s), gave birth to the personal computer. For a history of the computer from the time when 'computers were people' until roughly the mid 1990s, see Campbell-Kelly & Aspray 1996. For an overarching treatment of computers and computer history, see Rojas (ed.) 2001.

metaphysical assumptions, the failure has to be addressed and must be taken into account. Assessing such a failure cannot simply be ironed out by pursuing the same theoretical path, but necessitates that one puts into question the very theoretical framework and its underlying assumptions.

As can be expected, it was outrageous that a philosopher not trained in technical work would suggest in advance that the research program in AI was doomed to failure. Seymour Papert, one of the co-founders of MIT's AI laboratory, as well as creator of the programming language LOGO, can be considered the first of Dreyfus's critics. He famously wrote in 1968 a lengthy memo, which bore the title *The Artificial Intelligence of Hubert Dreyfus. A Budget of Fallacies*, dealing with Dreyfus's report on the advances of AI. According to Dreyfus's later recollections, his report for the RAND Corporation concerning the state of the art of AI research, *Alchemy and Artificial Intelligence* (1965), was described by Papert as "sinister, dishonest, hilariously funny, and an incredible misrepresentation of history" (1992, p. 87). In this report—bearing a disdainful title for its overt reference to alchemy—Dreyfus confirmed for the first time what seemed to be a pattern in AI research and first attempts at implementation: early successful results accompanied with unmistakable signs of stagnation (1965, pp. 9-17). The report develops the topics which Dreyfus would deepen later on in his career, but what probably caused all the fuss was the suggestion that an information processing device might and, indeed, does differ over the way a human agent thinks or 'processes information' (1965, pp. 18-46). According to Dreyfus, there were certain underlying assumptions taken for granted by AI researchers (such as the *petitio principii*, typical of cognitivism, according to which human agents actually 'process information') which hid a series of theoretical difficulties that were not being properly addressed. These misconceptions, in Dreyfus's own words, masked the seriousness of current difficulties (1965, pp. 46-64).

Papert's immediate reaction to Dreyfus was of ethical nature. According to Papert, the report was irresponsible because Dreyfus's facts "are almost always wrong; his insight into programming is so poor that he classifies as impossible programs a beginner could write" (1968, p. 2). Moreover, Dreyfus's report is full with "technical nonsense that pervades every paragraph" (*idem*), and this because he "knows nothing about the technical issues and barely understands the language used" (1968, p. 7). Papert's sense of outrage further increases in observing that "much of Dreyfus's 'penetrating analysis' (as A. Oettinger has called it) is generated by collecting specific difficulties Simon and Newell

report as *technical* problems for *particular* programs, and simply declaring them to be absolute obstacles for *all possible programs*" (1968, p. 9). Papert even thinks this deserves a meditation on the place of the humanities in higher education institutions:

> I sympathise with 'humanists' who fear that technical developments threaten our social structure, our traditional image of ourselves and our cultural values. But there is a vastly greater danger in abandoning the tradition of intellectually responsible and informed inquiry in the futile hope of an easy resolution of theses conflicts. The steady encroachment of the computer must be *faced*. It is cowardice to respond by filling humanities departments with 'phenomenologists' who assure us that the computer is barred by its finite number of states from encroaching further into the areas of activity they regard as 'uniquely human'. (1968, p. 3)

In a nutshell, Dreyfus lacks, according to Papert, not only the technical competence that would make him a serious and fruitful critic of AI, but above all he lacks academic integrity. "Our culture is indeed in a desperately critical condition if its values must be defended by allowing muddled thinking to depose academic integrity" (*idem*), adds Papert. And this explains very well how much Papert values those 'humanists', who he hastens to always put in quotation marks in his memo. After all, Dreyfus's arguments "must be read as literary conceits with deep 'humanist' content" (1968, p. 3). 'Humanism', in this special Papertian sense ascribed to Dreyfus (and apparently also to 'phenomenologists' filling humanities departments!), must come to grips with its romantic, crypto-theological, image of human being, along with its exaggeratedly simplified image of the computer, which "leaves the layman aghast at the suggestion that a robot could take dictation as well as a secretary" (1968, p. 7).

As is widely known, this sort of reactions had an immediate bearing on Dreyfus's intellectual honesty being put into question and his profession at the university being threatened. And understandably so, for Dreyfus's opinions on the matter[10] can arguably be recalled as giving birth to the

[10] In a pre-print version of a Dreyfus paper (2007) which circulated online (and still does: see URL: <http://leidlmair.at/doc/WhyHeideggerianAIFailed.pdf>, retrieved: August 20th, 2012), Dreyfus credits his work as affecting negatively the MIT AI Laboratory budget: "After I published, *What Computers Can't Do* in l972 and pointed out this difficulty among many others, my MIT computer colleagues, rather than facing my criticism, tried to keep me from getting tenure on the grounds that my affiliation with MIT would give undeserved credibility to my 'fallacies', and so would prevent the AI Lab from continuing to receive research grants from the Defense Department. The AI researchers were right to worry. I was considering hiring an actor to impersonate an officer from DARPA and to be seen having lunch with him at the MIT Faculty Club. (A plan cut short when Jerry Wiesner, the President of MIT, after consulting with Harvard and Russian computer scientists and himself reading my book, personally granted me tenure.) I did, however, later get called to Washington by DARPA to give my views, and the AI Lab did loose DARPA support during what has come to be called the AI Winter." See note 7 of that manuscript, pp. 31-32.

so-called 'AI winter', that is, the period of reduced funding and increasing lack of interest in AI research.[11]

But all this requires further comments. Which expectations in early AI research were actually not fulfilled? Herbert Simon and Allen Newell, both founding members of the AI research program who were present at the Dartmouth conference in 1956, bethought of themselves as technological prophets.[12] Paying attention to what they thought was to be expected from AI research, helps in catching a first glimpse of theoretical pitfalls to come, for such wanton predictions would come to be hard to defend in the near future:

> It is not my aim to surprise or shock you if indeed that were possible in an age of nuclear fission and prospective interplanetary travel. But the simplest way I can summarize the situation is to say that there are now in the world machines that think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until—in a visible future—the range of problems they can handle will be coextensive with the range to which the human mind has been applied. (1958, p. 8)

Perhaps nobody has to remind us today (more than five decades after these somewhat extravagant declarations) of the fact that until now computers do not think—unless one entertained, of course, a petty and quite reductive understanding of human thinking. There are today, simply put, no computers that think, that learn, and that create in the same manner human agents are capable of those actions. Simon and Newell's expectations seem, from our present point of view, quite exaggerated, to say the least. However, if this was supposedly the state of intelligent machines in 1958, it is to be assumed that futurist predictions would not be less flamboyant. The paper just quoted by Simon and Newell serves to introduce 'heuristics' as the theoretical discipline that would permit simulating solving-problem

---

[11] AI winter consisted of several phases, among which the following are to be recalled: 1966: the definitive acceptance of the failure of machine translation. 1970: the abandonment of connectionism (that is, modeling of mental and behavioral phenomena as the emergent processes of interconnected networks of simple units). 1971-1975: DARPA's frustration with the Speech Understanding Research Program at Carnegie Mellon University. 1973: the negative effects of professor James Lighthill's report (Lighthill 1973), whose prognosis largely cast doubt on AI research in the United Kingdom. 1973-1974: DARPA's cutbacks to academic AI research in general. 1987: the collapse of the LISP machine market. 1988: the cancellation of new spending on AI projects by the Strategic Computing Initiative. 1993: expert systems slowly reaching the bottom. Throughout the 1990s: the quiet disappearance of the goals originally dreamed of by the Fifth-Generation Computer Systems project (an initiative by Japan's Ministry of International Trade and Industry), and the increasing bad reputation of AI as science. On AI winter, see Hendler (2008). On AI as an industrial and entrepreneurial failure, see Phillips (1999).

[12] The Dartmouth Summer Research Conference on Artificial Intelligence is often considered the seminal event for AI as a field. Besides Simon and Newell were also present John McCarthy, Marvin Minsky and Claude Shannon, among others. It was organized by McCarthy, who is famously credited for coining the term 'Artificial Intelligence'. On this event, see McCarthy's research proposal (1955).

processes which were allegedly carried out at the basic level of human intelligence. In the context of introducing heuristics, Simon and Newell (1958) were willing to make the following predictions, that were to be realized within the next decade:

> · That within ten years a digital computer will be the world's chess champion, unless the rules bar it from competition.
>
> · That within ten years a digital computer will discover and prove an important new mathematical theorem.
>
> · That within ten years a digital computer will write music that will be accepted by critics as possessing considerable aesthetic value.
>
> · That within ten years most theories in psychology will take the form of computer programs, or of qualitative statements about the characteristics of computer programs. (pp. 7-8)

Only the fourth prediction, that psychology would become computational, did realize and became a sort of 'cultural common sense' on intelligence. And Deep Blue, the IBM supercomputer, actually defeated World Chess Champion Garry Kasparov in 1997, but it is quite accepted now that whatever calculations were needed by the machine to anticipate all possible moves by its opponent, the program does not resemble, however, the way the human mind of a chest player works. Even Simon, who staunchly adheres to traditional AI, agrees that a machine playing chess at a master's level and winning by pure brute computational force, would not exhibit what AI researchers are ultimately looking for: "it would be Artificial Intelligence, but not cognitive science. Take the best chess programs... It does not tell anything about how a chess grandmaster thinks, or very little" (1995, p. 243). And cognitive science was indeed the supreme aspiration of AI.

Why were these expectations not weeded out as laughable and illusory (and this precisely at a time when computers had less power than today's pocket calculators)?[13] Why was so much expected from AI as a research program? It was not only epistemological optimism with no basis, but above all the unchampioned assumption of a series of philosophical tenets. Particularly, there were four assumptions underlying persistent optimism: the so-called biological, psychological, epistemological, and ontological assumptions (Dreyfus 1992, pp. 153-227). In order to discuss these unrivaled

[13] No one surely needs hesitate to admit that a calculator "'has no idea' what numbers are. It can't count; it can't explain anything it does; it can't tell a proper fraction from a buffalo chip; and it doesn't care. *All* it can do is crank through four mindless algorithms, depending on which buttons are pushed. A number of considerably more elaborate systems, such as automatic bank tellers, word processors, and many computer games, belong in essentially the same category" (Haugeland 1989, p. 122). As it seems, it did not seem at all ludicrous to maintain quite the contrary at the time of the launching of the AI research program.

assumptions, a careful perusal of two classical papers by Simon and Newell will now be most fitting: the one that has been quoted already, "Heuristic Problem Solving: The Next Advance in Operations Research" (1958) and "Computer Simulation of Human Thinking and Problem Solving" (1962).

The paper on heuristics, as a programatic proposal, was written by both Simon and Newell, but presented by Simon alone as keynote speech at the banquet of the Twelfth National Meeting of the Operations Research Society of America, Pittsburgh, in 1957. Simon, a 1978 Nobel laureate trained in economy and political science, has been very influential in the sociological study of entrepreneurial organizations.[14] Moreover, advances on operations research, within which heuristics was supposed to play a decisive role, are tainted in Simon's work with an entrepreneurial and industrial aura. In addition, it is worth mentioning that, as part of the paper's argument, Simon and Newell understood their own theoretical project as situated historically in conjunction with Charles Babbage (1791-1871), "patron saint of our profession" (1958, p. 1). The French mathematician and engineer Gaspard de Prony (1755-1839) is also an important figure, given that his logarithmic and trigonometric tables suggested to Babbage "that machinery could replace human labor in the clerical phases of the task, and that started him on the undertaking of designing and constructing an automatic calculating engine" (Simon & Newell 1958, p. 3). According to Simon and Newell, the promise of designing such an engine amounts to the invention of a mathematical mechanism that could solve human problems, which could now be defined unambiguously in terms of how they are structured. In this sense, a well-structured problem must satisfy the following criteria:

> · It can be described in terms of numerical variables, scalar and vector quantities.
>
> · The goals to be attained can be specified in terms of a well-defined objective function—for example, the maximization of profit or the minimization of cost.
>
> · There exist computational routines (algorithms) that permit the solution to be found and stated in actual numerical terms. Common examples of such algorithms which have played an important role in operations research are maximization procedures in the calculus and calculus of variations, linear-programming algorithms like the stepping-stone and simplex methods, Monte Carlo techniques, and so on. (1958, pp. 4-5)[15]

[14] Some essays on economical rationality and social planning are to be found in Simon (1996).

[15] For a careful explanation of these algorithmic techniques and methods as quantitative approaches to decision making, which is was Simon and Newell had in mind, see Anderson et al. (2008), mainly chapter 17 ff.

In sum: "well-structured problems are those that can be formulated explicitly and quantitatively, and that can then be solved by known and feasible computational techniques" (1958, p. 5). Thus, there is a certain sense of distaste in Simon and Newell when it comes to assessing innumerable situations for which there exist no numerical variables but only "symbolic or verbal" and, that means, "vague and nonquantitative" (*idem*) ones. Simon and Newell's negative appraisal of this situation is to be expected: "there are many practical problems—it would be accurate to say 'most practical problems'—for which computational algorithms simply are not available" (*idem*). In the absence of computational algorithms, one encounters ill-structured problems which cannot be quantified and, consequently, must belong to the province of "judgement and intuition", for they pertain more often to "a matter of hunch than of calculation" (*idem*). It is for this reason that heuristics is proposed as a method in order to get rid of ill-structured problems—a sort of mechanics suitable for practical judgement in everyday life situations: "In dealing with the ill-structured problems of management we have not had the mathematical tools we have needed—we have not had 'judgement mechanics' to match quantum mechanics" (1958, p. 6).

But, according to Simon and Newell's optimism nonpareil,

> we now have the elements of a theory of heuristic (as contrasted with algorithmic) problem solving; and we can use this theory both to understand human heuristic processes and to simulate such processes with digital computers. Intuition, insight, and learning are no longer exclusive possessions of humans: any large high-speed computer can be programmed to exhibit them also. (*idem*)

Heuristics is the result of a series of investigations carried out by Simon, Newell and J. C. Shaw for the RAND Corporation during the 1950s and 1960s which were meant to give the computer the ability "to discover proofs for mathematical theorems—not to verify proofs, it should be noted, for a simple algorithm could be devised for that, but to perform the 'creative' and 'intuitive' activities of a scientist seeking the proof of a theorem" (1958, p. 7). There can be little doubt that the project was thus peppered with a grandiose philosophical task: "The research on heuristic problem solving will be applied to understanding the human mind. With the aid of heuristic programs, we will help man obey the ancient injunction: Know thyself. And knowing himself, he may learn to use advances of knowledge to benefit, rather than destroy, the human species" (1958, p. 8).

For the aforementioned reasons, it is easy to conclude that these pioneering AI researchers were not at work in order to obtain partial—or merely engineering—results. The project was meant to

simulate human intelligence and all its attributes, even to artificially reproduce the human mind. There is no inappropriate meddling of philosophy in engineering (as suggested in Papert's memo, 1968), because the ontological objective to be attained was quite clear: it was not only about implementing a series of technological mechanisms, but above all about simulating, and thus comprehending thoroughly, the essence of proto-operative subjectivity. If AI is philosophically interesting at all, it has to do with the fact that a group of well-funded researchers (with strong connections with the government of the United States, military agencies, and industry) had been trying to turn modern rationalism into a technological research program. The substantial impact of heuristics problem solving on research to be expected displayed a very 'scientistic' outlook[16] for future implementation, and should have certainly whetted the curiosity of everyone present in Simon's public address: "[w]hen machines will have minds, we can create copies of these minds as cheaply as we can now print books" (1958, p. 9).

Now, within a historical perspective, 'strong AI' is to be taken as a belated characterization of the project according to which the original purpose of programming intelligence was not realizable at all. (A great deal of work done in AI these days has abandoned its philosophical outlook and focuses exclusively on engineering tasks.) But it ought to be noted that the project of simulating proto-operative subjectivity, that which constitutes the workings of a human mind, or cognition, did not owe its implementation failure to the absence at the time of better theories for new mechanisms. The very conception of mind being entertained permits one to suspect of a 'chronicle of a death foretold'—as Gabriel García Marquez' famous short story title goes.

In "Computer Simulation of Human Thinking and Problem Solving" (1962), Simon and Newell presented a system they named GPS: General Problem Solver. The paper commences by announcing what seemed to be obvious about computers at the time:

> It is no longer necessary to argue that computers can be used to simulate human thinking or to explain in general terms how such simulation can be carried out. A dozen or more programs have been written and tested that perform some of the interesting symbol-manipulating, problem-solving tasks that human beings can perform and that do so in a manner which simulates, at least

[16] 'Scientism' can be defined as the attempt to extend the research methods of the natural sciences to every field of human knowledge. This endeavor is made in order to reach a 'Modern Synthesis' (see Wilson 1978, p. 90). According to Stenmark (1997), scientistic attitudes not only seek a scathing reduction of the humanities and social sciences to biology, "but biology is also reduced to chemistry, and chemistry to physics" (p. 16). In this regard, the reduction of mentality to neurophysiology or computation is all the more scientistic. On scientism, see Sorell 1991 and Putnam 1992.

> in some general respects, the way in which humans do these tasks. Computer programs can now play chess and checkers, find proofs for theorems in geometry and logic, compose music, balance assembly lines, design electric motor and generators, memorize nonsense syllables, form concepts, and learn to read. (1962, p. 137)

In accordance with its creators, GPS could be defined "as a system of methods—believed to be those commonly possessed by intelligent college students—that turn out to be helpful in many situations where a person confronts problems for which he does not possess special methods of attack" (1962, p. 138). This is why, when a person has to tackle a problem, she follows strictly a succession of unconscious rules, and this unconscious mechanism is always working every time alternatives for the resolution of difficulties have to be elucidated. At the hypothetical level, heuristics would allow for the algorithmic regimentation of the solving process commonly known as 'thinking'.

The feasibility of GPS assumes a basic philosophical tenet: there is, in principle, a subagential structure of how human beings think which is primarily constituted by a logical system of rules. But this is—it can be asserted pretty straightforwardly—a central philosophical conception of analytic philosophy. The Oxonian philosopher Peter Strawson has remarkably stated that there is an analogy to be made between human thought and a sort of grammar of thought, a sort of underlying structure of rules: "just as the grammarian, and especially the model modern grammarian, labours to produce a systematic account of the structure of rules which we effortless observe in speaking grammatically, so the philosopher labours to produce a systematic account of the general conceptual structure of which our daily practice shows us to have a tacit and unconscious mastery" (1992, p. 7). Pioneering AI researchers were laboring as well as analytic philosophers in that they were not just proposing and discussing new theories, but more than anything, getting underway the task of realizing technologically the philosophical assumption according to which there was, indeed, a systematic conceptual structure and a subagential level of human thought performance. Moreover, what was tacit and unconscious was that which, by means of certain algorithms assigned to a machine as a program, could in fact simulate the heuristic processes which would purportedly constitute a human mind. It should be noted that this tacit, underlying structure is also assumed in GPS:

> As a theory of human problem solving, GPS asserts that college students solve problems—at least problems of the sort for which the program has been tested—by carrying out this kind of organized ends-means analysis. It does not assert that the process is carried out consciously—it is easy to show that many steps in the problem-solving process do not reach conscious awareness. Nor does the theory assert that the process will appear particularly orderly to an observer who

> does not know the program detail or, for that matter, to the problem solver himself. It does assert that, if we compare that part of the human subject's problem-solving behavior which we can observe—the steps he takes, his verbalizations—with the process carried out by the computer, they will be substantially the same. (1962, p. 141)

This is what clearly constitutes a common ground between Simon, Newell and Strawson: in that they assert that there is an unconscious process, a mainly subagential third-person perspective, capable of sustaining all human conscious coping, that is, the agent's being-in-the-world. The point is of course to demonstrate that, as a matter of fact, AI is able to execute this assumption on practical ground. Should this assumption be true, even at a reduced scale, that is to say, if it were, in effect, a fact that human agents make use of heuristics in problem-solving processes and that these can be programmed technologically, then further investigation could well congratulate itself on a major success. Subsequent and more exciting results could thus be expected and the whole range of programming cognition would be reachable at the giddy pace of progress to which technological commodities have us accustomed.

Dreyfus's early but firm nay-saying appears today—four decades after his highly critical lunge against the false expectations of AI researchers—as a daring insolence. Shouldn't he have waited a few decades to at least acquire a little perspective on the development of more powerful machines fitted for the magnificent task of programming human thinking? Isn't it precisely because of the vertiginous pace in which technology develops that we err most of the time and our own predictions tend to fall short? As Crevier (1993) remarks referring to Dreyfus's early memo against AI: "it was almost uncanny for a non-expert in computer science to anticipate as early as 1965 the difficulties AI would run into, and to point out why!" (p. 125). But if Dreyfus was right even from the very beginning, it was on the basis of having determined a very abstract philosophical conception, nonembedded and nonembodied in nature, an empirical metaphysics, being held without question in AI. This was indeed an analytical hypothesis of metaphysical dimensions, being nonetheless held as a mere platitude, according to which the assignation of rules to a machine about atomic facts could result in reproducing the totality of human intelligent behavior. In this vein, the world is nothing but a collection of objects with properties and the resulting processes of those fixed objectified things.

This somewhat naïve, and certainly nonphenomenological characterization of the world, serves Dreyfus as motivation for exercising a critique of the limits of AI. In the latest edition of *What*

Computers Still Can't Do (1992), the former subtitle has been modified: *Critique of Artificial Reason*. This critique intended to be phenomenological and thus Papert was right when fearing (1968, p. 3) that phenomenology would have something critical to say about AI as a whole. But he was mistaken by supposing that this critique was to be easily deemed 'romantic'—having simply to do with a somewhat theological image of human being or, more precisely, with the fears accompanying the social and cultural changes being brought about by the information and communication technologies of the time. For reasons that will be explained in greater detail further below, Dreyfus has been surely very important for the reception of phenomenology in the—customarily—analytically dominated departments of philosophy in the United States of America. But phenomenology, as a way of practicing philosophy, has gradually entered the stage in the English speaking world not without basic misunderstandings which are present even today. And Dreyfus has influenced a series of researchers in AI who have tried to cope with this phenomenological heritage.

Now, however, before moving on to the substance of the breakthrough of phenomenology in AI and cognitive science and going into the heart of the matter staged in this investigation, a definition of what Dreyfus first criticized—the foundational approach to achieving AI—is needed.

GOFAI as a Degenerating Program

Haugeland famously christened the oldest approach to achieving AI 'Good Old Fashioned Artificial Intelligence'—or GOFAI, for short. What is crucial in defining what GOFAI amounts to, as a branch of cognitive science, "rests on a particular theory of intelligence and thought—essentially Hobbes's idea that ratiocination is computation" (Haugeland 1989, p. 112). As is widely known, Hobbes clung harshly to the view that thought, that is, reason as a whole, could be reduced to calculation:

> Out of which we may define (that is to say determine) that which is meant by this word *reason* when we reckon it amongst the faculties of the mind. For REASON, in this sense, is nothing but *reckoning* (that is, adding and subtracting) of the consequences of general names agreed upon for the *marking* and *signifying* of our thoughts; I say *marking* them, when we reckon by ourselves; and *signifying,* when we demonstrate or approve our reckoning to other men. (2005, p. 34)[17]

[17] As near as I can tell, however, it is but Leibniz to whom the exhortation *calculemus!*—directed to his objectors against the backdrop of scorn being heaped upon his dream of a calculus machine—is often attributed. With regard to this chapter in the history of computational machinery, see Stein et al. (2006). On Hobbes's *Leviathan* and machines, see Dyson (2007), pp. 1-13.

Insofar as GOFAI remains faithful to the Hobbesian heritage, at least two claims are essential to this approach:

> · Our ability to deal with things intelligently is due to our capacity to think about them reasonably (including subconscious thinking).
>
> · Our capacity to think about things reasonably amounts to a faculty for internal 'automatic' symbol manipulation. (Haugeland 1989, p. 113)

By sifting through these two claims, it now turns out to be apparent that Simon and Newell were hard at work attempting to put them into some sort of practice.

Dreyfus analysis of the then current research in AI spans a decade (1957-1967) in which the foundational approaches of GOFAI (Cognitive Simulation—1957-1962—and Semantic Information Processing—1962-1967) were put into work. Now, his first critique of traditional AI or GOFAI focused primarily on showing the view that the attempt to turn rationalism into a technological research program was, indeed, a dubious slippery slope: "I began to suspect that… by combining representationalism, conceptualism, formalism, and logical atomism into a research program, AI researchers had condemned their enterprise to reenact a failure" (Dreyfus 2007, pp. 247-248). It bears some thought just to be reminded of the fact that this suspicion was first formulated four decades ago, which, from today's point of view, would seem like a very daring wager. However, Dreyfus's confidence was sound from the beginning, since his search for signs showing that the whole AI research program was degenerating was based on his certainty that he was refuting a particular philosophical conception underlying the program: he was attacking the framework itself, not particular experimental results. Not even Moore's law (2005)—according to which over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every two years—would prevent Dreyfus of his staunch criticisms, for predictions based on false or one-sided philosophical assumptions can in advance be relegated to the trash can of scientific illusions.

That things were not going on well, or at least not as well as expected, was agreed somewhat timidly by some researchers who named purported difficulties for particular programs coming across and wreaking havoc, which is normal in scientific research. Marvin Minsky, at the time codirector of the AI research laboratory at MIT, even noticing the problems deriving from GPS and heuristics, "was convinced that representing a few million facts about objects including their functions, would solve

what had come to be called the commonsense knowledge problem" (Dreyfus 2007, p. 248). Partly inspired by Minsky, John McCarthy famously wrote a MIT lab report entitled *Programs with Common Sense* (1958) in which he envisaged the possibility of writing a program ('Advice Taker') that would "manipulate in a suitable formal language (most likely a part of the predicate calculus) common instrumental statements" (p. 1). A property would make available to 'Advice Taker' a wide class of immediate logical consequences of anything it is told and this property "is expected to have much in common with what makes us describe certain humans as having common sense" (1958, p. 2). A program has common sense, adds McCarthy, "if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows" (*idem*).

Thus, the first obstacle to simulating cognition technologically was no other than common sense, even it being 'the most fairly distributed thing in the world'—as Descartes once remarked. But, as AI researchers were beginning to find out, giving an account of the elegance and simplicity of the way human beings cope with the world, with little artifice and little effort, that is, paying attention just to those aspects that matter while at the same time ignoring, or assuming unproblematically, nonimportant ones, is the most difficult of tasks. As a matter of fact, when Minsky was asked at *Wired* magazine why such a high profile character of AI declared it to be brain-dead since the 1970s, he then answered: "[t]here is no computer that has common sense" (2012). The recognition of this failure goes back to the shortcomings brought about by the frame problem,[18] which is at the same time inextricably linked with the problem of commonsense knowledge.

The frame problem has to do with contextual relevance in a giving situation in which typically human agents cope with certain situations in the world and the way a program might be designed to actually be endowed with this kind of skillful coping in situation-bound contextuality. Shanahan (2004) has put it this way: "how do we account for the apparent ability to make decisions on the basis only of what is relevant to an ongoing situation without having explicitly to consider all that is not relevant?" A program might be written with propositions about objects and their properties about 'the external

[18] Dreyfus's interest in the frame problem was mainly ontological, without eluding the epistemological difficulties deriving therefrom. But the frame problem made its first appearance in logic. It was originally presented as a problem in a McCarthy and Hayes paper (1969). The ensuing question was raised in this context: "Using mathematical logic, how is it possible to write formulae that describe the effects of actions without having to write a large number of accompanying formulae that describe the mundane, obvious non-effects of those actions?" (Shanahan 2004, p. 2). Shanahan has argued, however, that the frame problem in its technical, logical, guise is more or less solved (1997).

world' but this brings to bear a number of distinct assumptions. In the first place, [1] that the world is something like a complex structure of fixed functional objects about which intelligent agents are conscious and attend to when coping with it, while ignoring some aspects of the situation not currently important; and, in the second place, [2] that an intelligent program simulating cognition must be endowed with a script (Schank & Abelson 1977), provided it details both what the program must consider attentively, as well as ignore, when coping with a common situation; whereupon 'knowing' for the program would be tantamount to being provided with a mental scale model, complete with automatic causal side effects (Haugeland 1996b, p. 92). Mental representation is therefore basic for a computational theory of mind, for such a theory supposes that cognitive states and processes are constituted by the occurrence, transformation, and storage in the mind/brain of information-bearing structures called representations (Pitt 2000). For Schank and Abelson, 'knowledge' is, then, processed in the form of planning scripts anticipating a number of distinct actions along with their non-effects:

> A script is a structure that describes appropriate sequences of events in a particular context. A script is made up of slots and requirements about what can fill those slots. The structure is an interconnected whole, and what is in one slot affects what can be in another. A script is a predetermined, stereotyped sequence of actions that defines a well-known situation. (Schank & Abelson 1977, p. 41)

Now, given that a few million facts about objects including their functions must be represented, a program purportedly provided with situation-bound knowledge must make constant reference to its recursive storage of concepts which represent symbolically the external world. It must be thus endowed with a very complete list of concepts and context-bound case scenarios. This means such a program encounters first and foremost the problem of context relevance and, actually, it clashes very violently with the task of determining what is at stake in a given context-bound situation. But this difficulty is, indeed, very abstract and theoretical for it is handsomely avoided by human agents. *The frame problem is not a human problem*. Quite conversely, in human coping, environmental changes are always variable and constant, whereby no change alters the big picture of what is being carried out, that is, no other basic environmental facts are thereby altered. It runs counter to experience to suppose that there is a structure of fixed objects with permanent functions being recursively anticipated every time. Moreover, the meaning of a particular action does not seem to be located unambiguously in a preset constituted

by fixed properties. Another way in which the frame problem can be formulated has been offered by Wheeler:

> Given a dynamically changing world, how is a nonmagical system… to take account of those state changes in that world (self-induced or otherwise) that matter, and those unchanged states in that world that matter, while ignoring those that do not? And how is that system to retrieve and (if necessary) to revise, out of all the beliefs that it possesses, just those beliefs that are relevant in some particular context of action? (2005, p. 179)

The world is dynamic, and if it is not magical—which seems to be the case taking account of what we already know from biology, chemistry, and physics—there should be a certain way of coping logically with it, abiding by the laws of nature. Hence, there should be a list of rules being followed whenever human agents cope with the world: there must be some sort of 'information processing'. Minsky, perhaps drawing on Schank and Abelson's scripts, postulated the need of a memorized framework describing in detail a number of situations constituted by constant changes in the environment: "[a] *frame* is a data structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party" (1974, p. 111-112). Moreover, several kinds of information are attached to each frame, some of it being about how to use the frame and about what one can expect to happen next. Thus, frames would anticipate situation-bound scenarios and pitfalls emerging therefrom in order to satisfactorily cope with what is being attended to and what need not be taken in consideration. As though all these aspects constituting the wide range of the frame problem were not enough, a theory of belief is also here afoot, because context sensitivity or situation-bound knowledge also entails the question about "*which* beliefs need to be considered, and which need not?" (Haugeland 1996b, p. 82). And this brings about another difficulty (which will be—phenomenologically—dealt with later on) that has to do with the question whether beliefs and, particularly, their specific propositional 'aboutness,' are original in embedded coping instead of sheerly derivative. For it may be suitable to underscore that perhaps "the frame problem may be an artifact of assuming that mental representation is quasi-linguistic" (Haugeland 1996b, p. 92), that is, essentially propositional and apophantic.

Be that as it may, from the point of view of computation (McDermott 1996), the frame problem necessitates to solve a technical shortcoming, as formulated by Shanahan (2004): "how to compute the consequences of an action without the computation having to range over the action's non-effects." In this regard, several strategies have been offered to solve the computational side of the

frame problem: mainly, the cheap-test and the sleeping-dog strategies (Haugeland 1996b). The cheap test strategy, on the one hand, deploys an effective list scanning that searches for that which is more relevant and remains unchanged. The strategy assumes that what is needed is "a prior *characterization* of events and facts, based on which types of events affect which types of facts" (Haugeland 1996b, p. 83). The sleeping dog strategy, on the other hand, is sensitive to the consequences deriving from the so-called 'common sense law of inertia,' to which Shanahan has dedicated an entire book (1997), in which he claims to have rendered a solution to the logical version of the frame problem. According to this 'common sense law of inertia,' the properties belonging to a specific situation can be assumed by default as fixed, so they do not change as the result of an action. To put it slightly different, "the sleeping dog strategy is to let everything lie, unless there's some positive reason not to" (Haugeland 1996b, p. 84). But do these strategies (look rapidly at lists, let everything not currently involved lie unaffected and dormant) solve the relevance problem?

Dreyfus is not convinced, for it seems obvious to him that any AI program using frames will be unavoidably caught up in an infinite regress. A program might be provided with a handful of context-bound instructions, even millions of them, but a frame is not embedded in a situation,

> so in order to identify the possibly relevant facts in the current situation one would need a frame for recognizing that situation, etc. […] Any AI program using frames was going to be caught in a regress of frames for recognizing relevant frames for recognizing relevant facts, and… therefore, the commonsense knowledge storage and retrieval problem wasn't just a problem; it was a sign that something was seriously wrong with the whole approach. (Dreyfus 2007, p. 248)

Given the wide problematical range of situation-bound contextuality, Minsky's optimism seems simply outrageous: "within a generation… the problem of creating 'artificial intelligence' will substantially be solved" (1967, p. 2). But this naïvety begins to cloud up when critical analysis brings the assumptions underlying traditional AI to the foreground. Dreyfus calls them the biological, psychological, epistemological and ontological assumptions underlying persistent optimism in AI.

Theoretical Assumptions Underlying GOFAI

The Biological Assumption

The biological assumption draws on a theory of information processing (Minsky ed., 1969), whereby the mind is conceived of as a computer. Two main premises, according to Scheutz, are buried in the computer metaphor of mind:

> · that the mind can somehow be understood as computation or be described by a program;
>
> · and that the same kind of relation that obtains between computational processes (i. e., executed programs) and computer hardware—the implementation relation—obtains between minds and brains, too. (2002, p. 12)

Dreyfus dismisses the metaphor abetted by computationalism and sharply criticizes it as "the naïve assumption that man is the walking example of a successful digital computer program" (1992, p. 159). But it is precisely under the auspices of computationalism, that traditional GOFAI programs such as heuristics and GPS begin to make sense. Perhaps it is true what Boden has asserted:

> to dismiss AI as philosophically bogus because of shortcomings in its earliest branch (GOFAI) would be like a seventeenth-century philosopher rejecting Galileo's suggestion that 'mathematics is the language of God' because—having no differential equations—he could not explain fluid dynamics. (1990, p. 19)

But it is nonetheless noteworthy that the computer metaphor of mind has not only caught on and remains indelibly fixed in the public's imagination, but remains as well a central tenet of great part of current cognitive science research. As a matter of fact, even today the orthodoxy in cognitive science research has it that general assumptions about the mind and intelligent thought and behavior are to be held, such as the following: the mind is an information processing system, a representational device and in some sense a computer.[19] According to Dreyfus, however, it suffices to discover that the brain processes information differently in order for this computational conception of mind—which pretends to be unambiguously 'biological'—to be shattered and consigned to oblivion.

Already in 1966, in a paper on cybernetics and the human brain, Walter Rosenblith argued that "detailed comparisons of the organization of computer systems and brains would prove equally frustrated and inconclusive" (*The American Scholar*, 274. Quoted by Dreyfus 1992, p. 162). Dreyfus presents opposing reasons to the assumption that biological systems can be best understood as information processors:

---

[19] See the preface to the Blackwell *Companion to Cognitive Science*, edited by Bechtel and Graham, 2008.

> The view that the brain as a general-purpose symbol-manipulating device operates like a digital computer is an empirical hypothesis which has had its day. No arguments as to the possibility of artificial intelligence can be drawn form current empirical evidence concerning the brain. In fact, the difference between the 'strongly interactive' nature of brain organization and the noninteractive character of machine organization suggests that insofar as arguments from biology are relevant, the evidence is against the possibility of using digital computers to produce intelligence. (1992, p. 162)

The Psychological Assumption

The question is whether cybernetics justifies the use of the computational metaphor of mind in psychology. As noted above, Simon and Newell (1958) had expected psychological theories of the future to become computational programs and this wholly in accord with a conception of AI as reverse engineering. Moreover, early AI researchers did not just wait for this to happen, but they began to write computer programs which simulated human processes, like Simon and Newell's joint efforts on heuristics and GPS. One enthusiast of the psychological assumption and early cognitive psychologist, the late Ulrich Neisser, contended that "the task of a psychologist trying to understand human cognition is analogous to that of a man trying to discover how a computer has been programmed" (*Cognitive Psychology*. Quoted by Dreyfus 1992, p. 164). The spectacular task of cognitive psychology would be, then, to discover the running program underlying the human mind.

That the mind 'processes information' is, nevertheless, very ambiguous, for what is to be understood by 'information' and 'processing'? In this regard, Dreyfus speaks of an improper argumentation in reasoning: "the fallacy of moving from the fact that the brain in some sense transforms its inputs to the conclusion that the brain or mind performs some sequence of discrete operations" (1992, p. 166). However, the concept of information remains unclear in that it admits of several interpretations. In a much-quoted passage, the 'father of information theory,' Claude Shannon, had already warned about not extrapolating the engineering aspects of information with those of semantic nature, because "these semantic aspects of communication are irrelevant to the engineering problem" (1948, p. 379). Warren Weaver explains Shannon's rejection of semantics in communication as follows:

> In fact, two messages, one of which is heavily loaded with meaning and the other of which is pure nonsense, can be exactly equivalent, from the present viewpoint, as regards information. It is this, undoubtedly, that Shannon means when he says the semantic aspects of communication are irrelevant to the engineering aspects. ('Recent Contributions to the Mathematical Theory of Communication.' Quoted by Dreyfus 1992, p. 165)

In this unjustified extrapolation, Dreyfus sees an illegitimate transformation of the mathematical theory of information into a theory of meaning: "Much of the literature of cognitive simulation gains its plausibility by shifting between the ordinary use of the term 'information' and the special technical sense the term has recently acquired" (1992, p. 166). Significantly though, Jerry Fodor assumes that the physical transactions that take place in the central nervous system "must satisfy such descriptions as 'monitoring texture gradients,' 'processing information about texture gradients,' 'computing derivatives of textures gradients,' etc." (1968, p. 632). But Dreyfus despairs of any account of this sort because for him a difference is to be made between a system being rule describable vs. rule governed. It is possible, of course, to describe the motion of the planets around the sun by rules, but it would be implausible to suggest that these rules actually govern their motion. To face this confusion, Dreyfus claims that "we need not conclude from the fact that all continuous physicochemical processes involved in human 'information processing' can in principle be formalized and calculated out discretely, that any discrete processes are actually taking place" (1992, p. 168).

The Epistemological Assumption

The powers of formalization are, nonetheless, very tempting, and on these the epistemological assumption bases his case. One must admit, though, that this assumption presents a great many difficulties which have to be borne carefully in mind. To this effect, the distinction between behavior describable by rules vs governed by those rules, might once again be useful in order to clarify the limits of formalization. While the planets, for example, are with most certainty not solving differential equations as they swing around the sun nor following any rules at all, "their behavior is nonetheless lawful, and to understand their behavior we find a formalism… which expresses their behavior as motion *according* to a rule" (Dreyfus 1992, p. 189). Although it is tempting to assume formalization can be generalized constituting a theory of knowledge, a confusion which emerges therefrom can be easily detected. Actually, formalization enables one to understand *competence*, what is being accomplished in skillful coping—riding a bike, say—inasmuch as competence can be formalized. But this possibility does not amount to an explanation of *performance*: "it tells us what it *is* to ride a bike successfully, but

nothing of what is going on in [an individual's] brain or in his mind when he performs the task" (Dreyfus 1992, p. 190). Consequently, Dreyfus identifies two axioms beloved of the epistemological assumption: [1] that all nonarbitrary behavior can be formalized and [2] that the formalism can be used to reproduce the behavior in question (*idem*). According to Dreyfus, these two claims emerge from an unjustified generalization that takes account of the success of physics, which therefore assumes there must be something like 'laws of behavior.'

The discussion whether or not there are such laws and what method, experimental or otherwise, should be employed in order to discover them, dates back to Hume's attempt to render a theory of human nature (2000) with the purpose of introducing—as the subtitle of the treatise states—'the experimental method of reasoning into moral subjects.' In this same vein, John Stuart Mill links the question whether moral sciences exist, or can exist with the possibility of laws of human behavior: "Are the actions of human beings, like all other natural events, subject to invariable laws?" (1994, p. 21). Mill undertakes the task of arguing at length that just as all natural phenomena are governed by universal laws, so the operations of men's minds are no exception. Since eventually all primary attendant circumstances of causality could be taken into account, one must conclude that human mind can be and should be a subject for scientific inquiry; or else, research should confine itself to mere desires and purposes, whose generalization would be tendentious at best. If we really want to calculate instead of merely having a hunch about issues concerning human actions (Simon & Newell 1958), such task of subsuming behavior under its laws should be at least tractable. Until that theory is available, following this reasoning, we will be stuck with no more than shaky intuitions.

Dreyfus thinks it both important and feasible to show that Allan Turing's ideas concerning the question *Can machines think?* (1950), are partly an inheritance of this discussion dealing with underlying laws of behavior, since there reigns as well the confusion according to which our ability to formalize competence amounts to a satisfactory explanation of performance (which is false). Indeed, Turing's problem is far more akin to Mill's problem than one would think at first glance. Turing conceives of the essence of digital computers in that they are rule-following machines (1950, p. 33) that are "intended to carry out any operations which could be done by a human computer" (1950, p. 32). It is interesting to note, nevertheless, that Turing finds the original question on the possibility of thinking machines "too

meaningless to deserve discussion" (1950, p. 38). For Turing, it should not worry anyone whether 'machines'—whatever the definition of machinery—can 'think'—whatever the definition of thought. In contradistinction with this way of asking the question, Turing is interested in knowing whether machinery can actually display some competence that would make one classify it as 'intelligent,' that is, as nondistinguishable from competences currently displayed by human agents. So when Minsky remarks "there is no reason to suppose that machines have any limitations not shared by man" (1967, p. vii), he is endorsing Turing's view that "at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (1950, p. 39).

But Turing is also sensitive to the confusion—which very often goes unnoticed—between 'rules of conduct' and 'laws of behavior':

> By 'rules of conduct' I mean precepts such as "Stop if you see read lights," on which one can act, and of which one can be conscious. By 'laws of behavior' I mean laws of nature as applied to a man's body such as "If you pinch him he will squeak." […] We believe that it is not only true that being regulated by laws of behavior implies being some sort of machine (though not necessarily a discrete state machine), but that conversely being such a machine implies being regulated by such laws. (1950, p. 48)

Dreyfus claims, however, that this is an unjustified generalization of Wittgenstein's argument that it is impossible to supply normative rules prescribing in advance the correct use of a word in all situations (1992, p. 192). Thus, Turing goes on to affirm roughly that although one cannot formulate, *de iure*, all normative rules that apply correctly to any given predicate, this does not speak against the possibility of formulating, *de facto*, those rules that describe how an agent applies such a predicate to a situation: "while Turing is ready to admit that it may in principle be impossible to provide a set of rules describing what a person *should* do in every circumstance, he holds there is no reason to doubt that one could in principle discover a set of rules describing what he *would* do" (Dreyfus 1992, p. 193). One could object to this that even if there is such a set of rules, this does not imply it can actually be formalized.

Turing concludes that there is no circumstance under which one could assert: "we have searched enough. There are no such laws" (1950, p. 49), but the ambiguity remains and both, the laws of behavior and of conduct, could in fact refer to meaningful human actions and to the physical regulations of an organism whatsoever. Given that human bodies are certainly part of the abstract

world described by physics, this very fact leads one to suppose that, as such physical objects, they must be rule-obeying. Regulations of such a physical object as the human body can be in principle formalized as the simple trajectory of a bullet or of a missile. So GOFAI brings forward the argument that the mind is a symbol-manipulating information-processing device which represents 'worldly facts' by recourse to logical operations. However, it could be a mistaken argument from the success of physics to assume that even motion in a physical body can be absolutely calculated and formalized. According to Bremmermann's limit, which calculates the maximum computational speed of a self-contained system in the material universe, this might well be physically impossible, because "no data processing system whether artificial or living can process more than 2 X 10^{24} bits per second per gram of its mass" (1962, p. 93).[20] For Dreyfus, this shows a special kind of impossibility in any attempt to simulate the brain as a physical system: "*the enormous calculations necessary may be precluded by the very laws of physics and information theory such calculations presuppose*" (Dreyfus 1992, p. 197).

So there seemed to be more factors at work than GOFAI researchers were able to reckon with, but these did not stop them from being tireless preachers of cognitive simulation. Yet, as Dreyfus noted,

> workers in the field of AI from Turing to Minsky seem to take refuge in this confusion between physical laws and information-processing rules to convince themselves that there is reason to suppose that human behavior can be formalized; that the burden of proof is on those who claim that "there are processes… which simply cannot be described in a formal language but which can nevertheless be carried out, e. g., by minds." (Minsky 1967, p. 107. Quoted by Dreyfus, *idem*)

The Ontological Assumption

In accordance with the ontological assumption, "everything essential to intelligent behavior must in principle be understandable in terms of a set of determinate independent elements" (Dreyfus 1992, p. 206). It needs scarcely be said that this assumption presupposes that the world is a collection of atomic facts, so knowledge about the world remains an immense bulk of discrete facts. This view is supported

[20] Bremermann's conjectures are, thus, the following: "There are π X 10^7 seconds in a year. The age of the earth is about 10^9 years, its mass less than 6 X 10^2 grams. Hence even a computer of the size of the earth could not process more than 10^{93} bits during a time equal to the age of the earth… Theorem proving and problem solving… lead to exponentially growing problem trees. If our conjecture is true then it seems that the difficulties that are currently encountered in the field of pattern recognition and theorem proving will not be resolved by sheer speed of data processing by some future super-computers" (1962, p. 94; quoted by Dreyfus 1992, p. 196).

by Minsky, who assesses common-every-day structures of knowledge to be quite large, provided they must include a collection of indispensable categories:

> Geometrical and mechanical properties of things and of space; uses and properties of a few thousand objects; hundreds of 'facts' about hundreds of people, thousands of facts about tens of people, tens of facts about thousands of people; hundreds of facts about hundreds of organizations. (*Semantic Information Processing*. Quoted by Dreyfus 1992, p. 209)

No wonder Minsky was enthusiastic about imparting 'context' to computers in the form of frames, and no wonder most hopes were based on the idea that, because the world consists in an immense set of discrete entities and their underlying processes, the problem of achieving AI is the problem of storing and accessing a large database. But just like with the frame problem being nonproblematic for intelligent agents embedded in coping with various situations, discreteness is once again a problem for computers, not for human agents. Moreover, entities are conceived of as discrete because they have to be specified *beforehand* by symbolic descriptions which must make plain and explicit what they are, but they do not appear discrete in coping, as phenomenological description would show.

Furthermore, conceiving of the world as a large collection of discrete entities is, when not controversial, a very clumsy and nonimaginative way of representing it. It is, at first sight, a simplification of situations and objects, for these appear intertwined with each other and they are meaningful just because they make reference to a wider context. In fact,

> to recognize an object as a chair, for example, means to understand its relation to other objects and to human beings. This involves a whole context of human activity of which the shape of our body, the institution of furniture, the inevitability of fatigue, constitute only a small part. And this factors in turn are no more isolable than the chair. They all may get their meaning in the context of human activity of which they form part. (Dreyfus 1992, p. 210)

For GOFAI it is of utmost importance that these many variable aspects involved in intelligent coping be made explicit. But can all situations of the world be explicitly specified? Of course, the answer to this question is a rather enthusiastic *yes* if one conceives of the world as a large collection of discrete objetive entities, and hence the early attempts to render 'micro-worlds' exhibiting the ontological assumption in small scale was now afoot.

A micro-world can be defined as "a contrived, artificial domain in which the possible objects, properties, and events are all narrowly and explicitly defined in advance" (Haugeland 1989: p. 185). Haugeland's own question with regard to micro-worlds is crucial: "Why bother with a fake little 'world'

like that?" (*idem*). Indeed, there are several reasons why GOFAI researchers expected so much from these contrived domains, somewhat pretentiously called 'worlds.'

Originally, Winograd (1972) erected the micro-world strategy in order to account in small scale for the variety of cognitive functions constituting natural language and understanding. If there is some place to start, it must be the simulation in micro-scale of worldly tasks and problems. Just like in other well-known attempts to build micro-worlds dealing with vision (Guzmán 1968; Waltz 1972), developmental intelligence (Minsky & Papert 1972), and learning and categorization (Winston 1975), the idea was to discover the most fundamental principles of cognition. Consequently, the strategy to go 'micro' in order to envisage in small scale the most basic aspects of understanding—whose complexity would later on be 'scaled up' in future programs—was due to the belief that it was feasible and desirable to abstract away the distracting complexities of the real world. Dreyfus was right in signaling how much GOFAI researchers drew from the success of physics, for this way of cutting through the distracting details of reality to expose the basic principles of cognition and understanding can be compared to the strategy of early physicists, who "ignored friction and deformation to get at the underlying universal laws of motion" (Haugeland 1989, p. 186). But by doing this—Dreyfus concludes—early AI researchers were condemning their enterprise to failure.

So the question has to be raised: can wordly experience be compared to a mental retrieval of a collection of facts about objects and their properties? More crucially: can the world and everything thereof be made explicit? Of course, the phenomenological influence (mostly from Heidegger and Merleau-Ponty) one can see working in Dreyfus's critical attitude towards GOFAI made him take issue with the ontological assumption, for it holds that the world is a collection of objects, and mind a database. From this follows that 'knowledge' is simply an interaction with instructions that have been made explicit (programmed) in advance. However, the world, says Dreyfus, is a background which is presupposed by all our knowledge, so it can never be made explicit because every description or representation of it will always presuppose it as the horizon which constitutes cognition. Or, to spell it out better: it is essential that the world remained implicit and nonthematized. The world is indeed concomitant of everything we do, but it does not make itself appear as object for theoretical inspection in the first place.

In the end, the conception of the world brought forward by GOFAI would seem at best a Pyrrhic victory: "the micro-worlds effort may be credited with showing that the world cannot be decomposed into independent fragments" (Haugeland 1989, p. 195). But can this discovery ignored by early AI researchers be empirically established? Was it even received and taken into account? Did early AI researchers learn something from these theoretical difficulties?

| 3 |

THE SCOPE OF DREYFUS'S CRITIQUE OF ARTIFICIAL REASON

Dreyfus's Alleged In-Principle Argument

The fact is, as has been exemplified by Papert's virulent memo, that Dreyfus's criticisms were, at best, received as the not quite informed intervention of a philosopher on a technical field he allegedly did not entirely understood.[21] At worst, mockery and plainly not taking account of them substituted any critical reflection regarding the limits and shortcomings AI was experimenting at the time.[22] Opinion is thus divided as to the influence that Dreyfus's critique of artificial reason has had and will have in the philosophy of cognitiva science.

There is nothing in the least odd that Crevier's treatment of this topic springs to mind as the typical attempt to reduce Dreyfus's position to criticisms which do not offer practicable solutions to the issues in AI he derided (see 1993, p. 132). The point is straightforward: if someone launches an aggressive attack against a research program, they must be able to show how an alternative explanation can serve to cause a substantial betterment of the theory being reviled. For Crevier, although some of the Dreyfusian criticisms did take on renewed currency later on, thus giving "credit to the perceptiveness of modern humanist philosophy, it is highly questionable whether they added up to the theoretical impossibility of artificial intelligence" (1993, p. 131). After all, the difficulties were being encountered by the researchers themselves working on the ground, so they were not suggested nor discovered by the philosopher waging the war.

---

[21] Claims abound as to how the philosopher has not been doing his homework of keeping up with the relevant scientific literature. See, for example, McCarthy's claim: "Hubert Dreyfus claims that 'symbolic AI' is a 'degenerating research program', i. e., is not making progress. It's hard to see how he would know, since he makes no claim to have read much of the recent literature" (1996, p. 143).

[22] Weizenbaum confessed to Crevier that within the AI community a conclusion was reached: "the best thing to do was to give Dreyfus the silent treatment. Just not to talk about him, not to try to defend against him, not to laugh at him, nothing. Basically, as far as the AI community was concerned, Dreyfus became a nonperson" (Crevier 1993, p. 123).

Strom and Darden speak of a long-term Dreyfusian influence on the field being probably minimal (1996, p. 151), but they concede as well that "there has been a fair amount of research done that is at least compatible with many of his essential views—enough to construct at least an outline of a research program" (1996, p. 169). Both these authors presented to several influential workers in AI the question of Dreyfus's contribution to cognitive science and the result was pretty mixed: David Israeli says no one he knew has been influenced by Dreyfus, while Terry Winograd thought virtually everyone had been influenced by him. Clancy was at least indirectly influenced by Dreyfus via Winograd, as was Brooks via Agre and Chapman (Brooks 1999a, p. 97). Dennett, who jokingly refers to Dreyfus as the 'guru of holism' (1998, p. 220), claims in an interview (1992) that his influence has been significant because, being right in some of his criticisms, AI researchers have discovered these problems on their own. Hence it seems needless to belabor that this controversy as to the scope of Dreyfus's influence in AI circles brings to the foreground the fact that his phenomenological critique of artificial reason has not received unanimous assent, but, as Ellrich (2003) suggests, it is at the very least considerably important.

In forthcoming chapters of this investigation, both the direct and indirect Dreyfusian influence on some field researchers will be dealt with in depth. Parallel research programs that can be found under the rubric of the embodied, embedded, extended, and enactive approaches (the so-called 4e approaches, for short) have not been necessarily influenced by Dreyfus's work, that is, at least not directly, albeit they certainly touch upon many of the issues raised earlier by Dreyfus himself and phenomenology in general. Many of the current discussions on AI, admits Koschmann somewhat reluctantly, "appear to call up themes... encountered in Dreyfus's earlier writings" (1996, p. 129).

However, what is striking is how the reception by field researchers has tended toward caricature and a manifest reduction to one single thesis: *Dreyfus thinks AI is impossible*. Putting it this way conceals the main philosophical Dreyfusian contributions to cognitive science. Haugeland has nicely summarized this state of affairs:

> Part I [of *What Computers Still Can't Do*], because it was the most combative (and also the easiest to understand), got most of the attention. Also, since that discussion was the most timely—hence the most quickly obsolete—it is what the excellent substantive introductions to the later editions have mainly brought up to date. An unfortunate consequence of these concentrations, however, is that the more interesting and enduring parts of the book, Parts II and III, have been somewhat eclipsed and even neglected. (1996c, p. 119)

Precisely those parts dealing with the surreptitious assumptions shoring up AI's optimism (Part II) and the ways offered by phenomenology to conceive of cognition away from those assumptions (Part III), are the ones constantly ignored by Dreyfus's critics.[23] In short, existential phenomenology, as practiced by Dreyfus, has been ignored *tout court*, or has been interpreted mistakenly as a sort of otherworldly, mystical touch or 'phenomenological feel' (Strom & Darden 1996, p. 161), which is even worse. This explains why reducing Dreyfus's work to one single thesis (not only that GOFAI as science appears to be for the time being in conceptual tatters, but foremost that AI is impossible in principle, that is, forever and ever), would seem to make a terrific case for a refutation.[24] If there is progress in any of the programs designed by GOFAI researchers, this would prove to a greater or lesser extent how Dreyfus's insistence on stagnation and degeneration is indeed mistaken and preposterous. His critique of artificial reason would be dogmatic.

So any notorious advances on chess programs, as is to be expected, would be a matter of great congratulation on many grounds for AI researchers. As a matter of fact, Strom and Darden (1996) specifically want to bring the chess case to the foreground in order to argue that traditional AI has risen to the challenge offered by Dreyfus. Let us remember that 1996, when Strom and Darden contributed a lengthy review on the third edition of Dreyfus's *magnum opus* for the journal *Artificial Intelligence*, was a year short of magnificent Deep Blue beating world chess champion Garry Kasparov. Chess progress was then exemplified, not by the IBM supercomputer, but by Deep Thought: saluted as "clearly a vindication for the traditional AI program of heuristic research" (Strom & Darden, p. 156). According to these critics, Dreyfus's argument against chess programs acquiring human performance of the game gravitates around the idea that human agents, when playing, deploy strategies which cannot be implemented by sheer brute computational force. Dreyfus has stressed that, if these programs are to exhibit enough prowess to match human performance at a master level, they must not figure out from scratch what to do each time. That would be, not only counterintuitive, but futile because, conversely, expert performance necessitates the sort of 'zeroing in' that depends upon fringe consciousness, ambiguity tolerance, essential vs inessential discrimination, and perspicuous grouping (Dreyfus 1992,

[23] Crevier speaks of those parts being "almost poetic" (1993, p. 129) that is, not of scientific value.

[24] Dreyfus's AI critique is often construed as a 'problem-in-principle' (see Crane 2003, p. 128).

pp. 120-128): all elements belonging uniquely to human agents, which avoid successfully the exhaustive search employed by computers.[25] The point is that Deep Thought does look up every possibility each time and actually figures out from scratch what to do next without zeroing in. How can it be so good a player, even with the dearth of intuitive, holistic possibilities available for human agents? According to Strom and Darden, whatever role these possibilities play in human expert performance, "Deep Thought's expert performance does not depend on the sort of holistic abilities Dreyfus attributes to people" (1996, p. 160). The machine need not imitate human performance but stick to results and emulate human competence.

Now, if this is the case with Deep Thought, which is "at least a 'good' player" (Strom & Darden 1996, *idem*), while being nonetheless easily defeated by Garry Kasparov and Michael Valvo, the triumph of Deep Blue over Kasparov—with its capability of evaluating 200 million positions per second—would seem to confirm that brute computational force can indeed achieve expert competence at least in some activities deserving the epithet of intelligent. In the aftermath of Deep Blue's victory, tainted with Kasparov's accusations that human chess players had intervened on behalf of the machine, thus cheating, Dreyfus's opinions on the matter were required by the press. On may 12, 1997 he appeared on a national public radio program (PBS *The NewsHour* with Jim Lehrer) along with Dennett as panel experts on the topic.[26] This discussion on the very possibility of machine intelligence shows the moot point between Dreyfus and Dennett over whether computation can be generalized to encompass all reality. Dreyfus, on the one hand, is onto something more important here than the mere rendering of a priori or in-principal arguments: the very difference between man and machine performance proves computer simulation wrong on fundamental grounds. Conversely, Dennett thinks the so-called unbridgeable gulf between man and machine is nothing other than false belief: "There is not any original, intrinsic intentionality. The intentionality that gets ascribed to complex intentional systems is all there is. It is an illusion that there is something more intrinsic or real" (1996, p. 66).[27]

---

[25] "In the first edition of this book I noted that good chess players don't seem to figure out from scratch what to do each time they make a move. Instead, they zero in on a certain aspect of the current position and figure out what to do from there" (Dreyfus 1992, xxviii).

[26] The transcript of the TV program can be viewed online: URL: <http://www.pbs.org/newshour/bb/entertainment/jan-june97/big_blue_5-12.html>. Retrieved: August 23, 2012.

[27] See also "Intentional Systems" in Dennett 1978.

Deep Blue's triumph over Kasparov or the 'brain's last stand'—as the event was then being depicted by the press—was quickly dismissed by Dreyfus, who thought it necessary to cast doubts as to its significance: "The reason the computer could win at chess (and everybody knew that eventually computers would win at chess) is because chess is a completely isolated domain. It doesn't connect up with the rest of human life, therefore, like arithmetic, it's completely formalizable, and you could, in principle, exhaust all the possibilities." What this shows, thinks Dreyfus, is that "in a world in which calculation is possible, brute force meaningless calculation, the computer will always beat people, but, in a world in which relevance and intelligence play a crucial role and meaning in concrete situations, the computer has always behaved miserably, and there's no reason to think that that will change with this victory." Dennett, true to his style, retorted that "the idea that there's something special about human intuition that is not capturable in the computer program is a sort of illusion, I think, when people talk about intuition. It's just because they don't know how something's done. If we didn't know how Deep Blue did what it did, we'd be very impressed with its intuitive powers, and we don't know how people live in the informal world very well."

One can swift through these claims a little further by taking a look at a dialogue (Dreyfus & Dennett 1997) on the same topic that Dreyfus and Dennett held on *Slate* as part of the E-Mail Debates of Newsworthy Topics section, which is a continuation of *The NewsHour* radio program.[28] In the exchange, Dennett reminds Dreyfus of a "tone of absolutism, as found in your book title... More specifically, in your insisting that these were not just the hard problems (you've been right about them being the tough problems all along) but being insoluble problems" (1997, p. 266). Dennett thinks this 'tone of absolutism' is most detrimental to Dreyfus's AI critique for its being based upon the sheer in-principle argument according to which brute computational force cannot be conceived of as the origin of intuition. At any rate, he credits Dreyfus with having a "knack for drawing a forbidding map of the hard problems that lie in the distant future" (1997, p. 276). Thus, Dennett and Dreyfus share the common assumption that symbolic AI is on the wrong track when attempting the grandiose task of producing cognition through artificial means, if this production amounts to imitating human

[28] See URL: <http://www.slate.com/articles/news_and_politics/dialogues/features/1997/artificial_intelligence/_2.html>. Retrieved: august 24th, 2012.

performance. They differ though over whether intuition, or protocol statements about machine traces —which is how Dennett (1968) framed the problem when he first criticized Dreyfus in the late sixties —is to be given the importance it receives in common-sensical parlance.

But Dreyfus and Dennett have been talking for years, so before staging the discussion between them which ranges from chess to robots, a detour leading to Dennett's earliest take on Dreyfus is afoot, for there is an interesting context of discussion belonging to this dialogue of four decades which deserves attention. In Dennett's very first publication, "Machine Traces and Protocol Statements" (1968), he advocates an incipient form of cognitivism, from which he has not backpedaled in his whole career and has rather reworked and clarified over the years. The paper's main argument smacks in fact of the notion of the disembodied brain in vat argument (see Putnam 1981, pp. 1-21) and of an understanding of computation as continuous with nature. If the connection between the possibility of disembodied cognition and the ultimate computational character of nature obtains, it would seem to follow that this amounts to the vindication of sheer computational force. Hence, the highly contentious (eliminativist and cognitivist) idea espoused by Dennett that there need not be a causal linkage between the computational workings of the mind (the machine traces) with the phenomenological experience deriving thereof (the protocol statements).[29]

But this is a dense problematic and some clarifications are thus in order. Dennett revises the central tenet regarding confirmation in simulation—whereupon simulation is to be understood as the aim of producing programs that solve problems in the same way people do—according to which a comparison must be sought between programs or machine traces and the protocol statements of a human subject.[30] According to this view, if there is a high degree of correspondence between the machine trace and the protocol statement, the exhibition of human intelligence can be properly accorded (see Dennett 1968, p. 155). However, Dennett's revision of this notion of correspondence amounts to its radical demise: "The absurdity lies in worrying about the discrepancy between machine

[29] This 'causal linkage' referred to here alludes to Putnam's argument that "there is no more reason to regard the machine's talk of apples as referring to real world apples than there is to regard the ant's 'drawing' as referring to Winston Churchill" (1981, p. 11). Even a computer excelling at the Turing test exhibits just a "syntactic play that *resembles* intelligent discourse" (*idem*), precisely because the causal linkage with the real world is lacking.

[30] This formulation of the tenet, as Dennett notes, was first exhibited by Feigenbaum and Feldman (see 1963, p. 3).

trace and protocol. Of course, human beings are information processors, and eventually we will find how, but not by slavishly trying to reproduce protocol statements" (1968, p 161).

This is the earliest formulation of Dennett's conception that phenomenology—in his view: the analysis of protocol statements as introspection data—is doomed to failure. Of course, the computer does not scan its own workings in order then to draw inferences about their interpretation, mainly because doing so does not explicate, say, how its printout comes about. It is only by looking at its workings as a programmed by-product of the very operations it is performing (Dennett 1968, p. 158) that 'intuition,' as contrasted with brute-force solution methods, shows its problematic face: "to speak of intuition is to deny that one knows how one arrived at the answer" (Dennett 1968, p. 159). Moreover, "intuition is not a name of a known or recognized means of processing information" (*idem*). Even if the brain operated on quite different principles than the machine, it does not follow that introspection brings forth something particularly interesting to the extant discussion. On this view, it would be preposterous to seek for a correspondence between a mental operation like addition and the way this very operation is introspected when protocol statements about it are required. Simply put: "a great deal of information must have been processed of which I can give no account in the protocol" (*idem*). We should not be bound by naïve intuition, for the idea that "the human protocol gives us valuable clues as to how human beings process information" (Dennett 1968, p. 160) must be abandoned. As public nay-saying against Dreyfus's *Alchemy and AI*, Dennett's 'Machine Traces and Protocol Statements' does not defer to phenomenology, whose objections to AI are nothing but a priori and, thus, dogmatic arguments.

Dennett has made this a central plank of his theory of mind, for herein resides his well-known patronizing contempt toward phenomenological research. Moreover, his anti-phenomenological stance (in favor of the computing machine and against the idol of human intuition) should be sound, inasmuch as it has been previously conjoined with the very principle of evolutionary theory. It is conspicuous that around the revolutionary thrust of Darwinism clusters a group of related ideas about life, the most important being Darwin's *strange inversion of reasoning*. Dennett understands this inversion to be a fundamental component of Darwin's dangerous idea (see Dennett 1995), which functions as

well as a theoretical justification of his continuous derogatory tone towards phenomenological research.

Let us look at this strage inversion of reasoning in more detail. The significance of Darwin's evolutionary theory has long been held hostage to all sorts of animosities stemming from its revolutionary character. It is worth recalling that pre-Darwinian thought clung harshly to the view of universal order from above. This conception of reality as a 'Great Chain of Being' (Lovejoy 1936) resorting to a *scala naturae*, which was entertained for about one thousand years, was crushed by evolutionary theory. Darwin's evolutionary twist consisted in touching the neuralgic nerve of the Great Chain of Being and showing why this traditional view of nature was wrongheaded. As McKenzie, one of Darwin's early critics, has very handsomely put it, Darwin, "by a strange inversion of reasoning, seems to think Absolute Ignorance fully qualified to take the place of Absolute Wisdom in all of the achievements of creative skill" (1868 as quoted by Dennett 2009, p. 10061). Not only an omniscient God was absent in Darwin's theory, but even a teleological ideal like Progress, which was added to the Great Chain of Being in the eighteenth century (Marks 2008, p. 69) was negligible. Conversely, it was Absolute Ignorance, that is, a mechanism of replication, a series of algorithms constituting 'natural selection,' the sole driving force for adaptive evolution, since natural selection can be understood as an algorithmic process (Dennett 1995, pp. 48 ff). In contrast to Absolute Wisdom being the *terminus a quo* of creation, Darwin's strange inversion of reasoning accounts for the not less odd fact that, whatever intelligence is, it is nothing other than a result, the *terminus ad quem*, of the evolutionary process. Intelligence is the result of ignorance. In fact, intelligence was preceded by a series of natural processes which were not intelligently designed beforehand, nor did they know what they were doing. These underlying processes are blind and unintuitive. Again, knowledge and intelligence are the result of evolution, not the other way around. So this approach brings exactly to the foreground McKenzie's insight on the significance of Darwin's theory: "In order to make a perfect and beautiful machine, it is not requisite to know how to make it" (quoted by Dennett 2009, p. 10061).

According to Dennett, what is most enticing about this whole issue is the extent to which Darwin's evolutionary theory can be imported and merged with computationalism. The central figure to the insight according to which computers are to be conceived of in conformity with Darwin's

strange inversion of reasoning is, of course, Turing. It is astonishing, in fact, to recall that computers were once large groups of people with degrees in mathematics doing what one might call 'clerkly work.' This is why the term 'computer' as we understand it today can be misleading for it actually defines the occupation of one who computes. It was thus needed that the ones doing the computing *knew* exactly what arithmetic was. However, the gist of Turing's own inversion of reasoning in computing was that this needed not be so. The 'clerks' doing the computing need not know what they are doing. As argued by Turing, a digital machine could be designed that could "mimic the actions of a human computer very closely" (1950, p. 34), that is, it could be endowed with a set of rules or a set of instructions. Dennett has summarized this idea as follows:

> Turing showed that it was possible to design machines... that were Absolutely Ignorant, but could do arithmetic perfectly. And he showed that, if they can do arithmetic, they can be given instructions in the impoverished terms that they do 'understand' that permits them to do anything computational. (2009, p. 10061)

What Turing demonstrated was the crucial idea that digital computers could exhibit all the prowess we now take for granted without knowing what they are doing. By being fully ignorant, they could produce knowledge. Paraphrasing (and inverting) McKenzie's assertion, as Dennett does, "in order to be a perfect and beautiful computing machine, it is not requisite to know what arithmetic is" (*idem*). In fact, many examples in nature itself of routines being carried out without any explicit comprehension of its rationale burgeon mightily. Dennett argues that "this is Turing's strange inversion of reasoning uncovered in nature" (2009, p. 10063). What the conjunction of evolutionary theory and computationalism teaches us about reality is that nature is computational *par excellence*. So whatever quizzical reserves one may still have regarding the full-blown consequences of the Darwinian revolution, it must be granted that deferring to it implies the primacy of a bubble-up theory as opposed to a trickle-down theory of creation, whose after-effect is detrimental, not only to the creationist view of nature as intelligent design, but to any primacy given to any trickle-down, 'intuitive' (phenomenological) process. According to Dennett, "it is undeniable that the other necessary competences of life are composable from unliving, uncomprehending parts; why should comprehension itself be the lone exception?" (2009, p. 10062). Just to nail down the point, Dennett concludes that those who cannot abide Darwin's strange inversion of reasoning—still looking, say, for irreducibly complex features or uncomputable aspects of human experience whatsoever—are to be

called mind creationists! (*idem*). Moreover, "what Darwin and Turing had both discovered, in their different ways, was the existence of competence without comprehension" (2012).

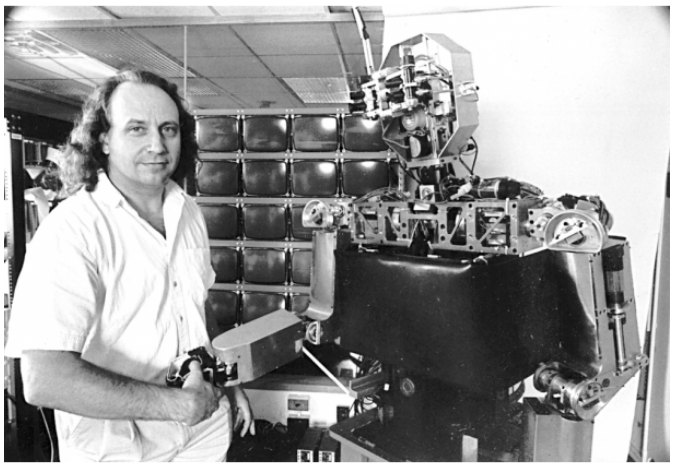
This brings us back to the difference concerning the original assessments of Deep Blue's victory pursued by Dennett and Dreyfus respectively. For the latter, although impressive, the win over Kasparov is no simulation of Kasparov's playing abilities, which can be confined to a few hundred plausible moves at most. Even one year before man's defeat to the processing power of the machine, Dreyfus submits accordingly that the difference as to how machines and human beings 'process information' makes the whole chess case uninteresting for cognitive science:

> if you had a machine playing chess by brute force, just simply counting out, it would still be Artificial Intelligence, but it certainly would be no contribution to cognitive science, because everyone knows that chess players could not possibly be processing information fast enough to be consciously or unconsciously counting out all possible positions up to seven-ply, which is something like twelve million moves, as good chess machines now do. (1996b, p. 74)

The idea that massive brute-force calculation can be construed as underlying a human agent's cognitive performance is what Dreyfus finds absurd. But is it? Dennett argues otherwise: "Kasparov's brain is a parallel-processing device composed of more than ten billion little robots. Neurons, like every other cell in a body, are robots, and the organized activity of ten billion little unthinking, uncomprehending robots *is* a form of brute-force computing, and surely intuition *is* nothing other than such an emergent product" (Dreyfus & Dennett 1997, p. 269). In the realm of symbolic AI reigns overt confusion precisely if (and when) Dreyfus's criticisms are taken seriously by GOFAI researchers, just as when a respected research program could bear the name "programs with common sense" (see McCarthy 1958) or when the frame-problem was recognized as such (see McCarthy & Hayes 1969). For Dennett, human being is a machine through and through and the puzzled sensation it gives us to think of consciousness as the still unconquered land for science, will probably feel—when it be theoretically deciphered—like "that mixture of amusement and letdown that often accompanies learning how a magic trick is done" (Dreyfus & Dennett 1997, p. 269).

In any case, for both Dreyfus and Dennett chess is certainly not the paradigmatic case that can advance the cause of AI. Far more interesting is the possibility of designing a humanoid robot, endowed with a more interesting competence than sheer counting out forcibly. The Cog Project

launched by the founder of the humanoid robotics group and the director of the MIT AI laboratory, Rodney Brooks, promised at the time (1993-2003) to build a humanoid robot on quite different theoretical principles than those of GOFAI. It would have to be flexible and embodied, so that it could interact online with its environment in order that, by doing so, it could develop similar kinds of representations to those of human agents.[31] Dennett, who was involved in the project as a consultant, was very excited with the idea of providing Cog with a humanoid infancy, by means of which it could interact with real people and learn to understand (see Dreyfus & Dennett 1997, p. 270). According to Dennett (1994), the practical requirements for making a conscious robot will rule out first of all an alleged special ingredient which is supposedly missing when trying to model cognition computationally. A direct disciple of Ryle—in whose very influential book, *The Concept of Mind* (1949), the derisive expression 'the ghost in the machine' first made its appearance—Dennett also thinks a theory of mind away from the 'Cartesian theater' is in much need to be devised. So both an *origin essentialism* (the thesis that brains can only be what they are, that is, they cannot be made of silicon, for example) and an *origin chauvinism* (the thesis that there is some mystic difference due simply to this very origin essentialism) have to be dismissed (see Dennett 1994, p. 136). After all, for Dennett, "an artificial brain is, on the face of it, as 'possible in principle' as an artificial heart, just much, much harder to make and hook up" (1994, p. 138).



Rodney Brooks with his humanoid robot Cog.

[31] Brooks' idea of robotics will be dealt with in depth in due course. See Brooks (1999) on the aims, scope and methodology of the Cog Project. The Cog Project website provided by MIT is still online. See URL: <http://www.ai.mit.edu/projects/humanoid-robotics-group/cog/cog.html>.

This is not to decry the difficulties involved in such an ambitious project. On the contrary, there is a practical reason to it, for "unless you saddle yourself with all the problems of making a concrete agent take care of itself in the real world, you will tend to overlook, underestimate, or misconstrue the deepest problems of design" (1994, p. 143). And besides, there is also no in-principal argument against artificiality—as the artificial heart example clearly shows. So "even if Cog really does have a *Lebenswelt*, it will not be the same as ours" (1994, p. 140).

Cog, finally, could be conceived of as a thought experiment (see Dennett 1997), but unlike philosophical thought experiments—where "the sun always shines, the batteries never go dead, and the actors and props always do exactly what philosophers' theories expect them to do" (1997, p. 252)—a project like Cog implies the possibility of going astray and committing theoretical howlers. This bequeaths the immediate suggestion that philosophy—"intellectual tennis without a net," memorable quote by Ronald Sousa brought up by Dennett (1997, p. 252)—would better work in close connection with engineering and robotics. Philosophy is just not part of those degenerating programs, whose dereliction Dreyfus is keen to criticize, because its hypothesis have not been tested on the grounds of technological implementation. What is exciting about AI is that we finally have the chance to put metaphysics to the empirical test.

Now, Cog is 'embodied' but frankly just to a certain extent. To begin with, it is not really made of the same relevant wetware our body is made of (see Dreyfus & Dennett 1997, p. 271), which allows for chemical changes due to hormones, adrenaline and the like. Neurophysiology holds that emotions depend on chemical processes. So how could Cog have any emotions? Moreover, is thinking a faculty wholly detached from an emotional stance? Dreyfus wryly retorts that this might be fundamental: "It may not be important that Cog's brain is silicon and ours protein, but it might be crucial that ours is wet and Cog's dry" (Dreyfus & Dennett 1997, p. 270). Dreyfus is specially baffled by Dennett's suggestion that Cog could "even exhibit a sense of humour" (Dennett 1994, p. 141). After all, the chemical effects affecting an agent's emotions and reactions are mediated by meaning. Hereto, thinks Dreyfus, is appended a series of serious considerations: how is Cog, a nonsocialized robot, to pick up emotions such as shame, guilt, and love from public narratives and exemplars? Which gender is Cog? How will the conversion from nonpropositional know-how into propositional knowing-that take place

in Cog? Aren't these fundamental questions that have to be addressed before embarking on the grandiose project of designing a humanoid robot? More puzzling is the attempt to imagine a humanoid robot changing its life abruptly, disclosing a new world in a particular way, by reinventing itself, by discovering its 'calling,' by embarking on an entirely new project or cultivating new interests, etc. (see Spinosa, Flores & Dreyfus 1997). Submitting this line of scathing criticism, Dreyfus thinks Dennett grossly underestimates how hard the problems are (Dreyfus & Dreyfus 1997, p. 274). But Dennett jokes about it: who cares if robots cannot disclose new worlds? "Well, so what? I didn't say you would want to marry one" (Dreyfus and Dennett 1997, p. 276).

Now, although Dreyfus and Dennett indulge in a certain agreement against GOFAI, because their approaches countenance the difficulties associated with it, it must be stressed that that is due to entirely different reasons. As is known, Dreyfus saluted the emergence of connectionism or parallel distributed processing as an alternative research project which could contribute to oust GOFAI from its pedestal. For Dreyfus, it still remains an open question "whether neural networks can be intelligent or whether network researchers, like AI researchers in the 1960s, are basing their hopes on *ad hoc* successes that may not be generalizable" (Dreyfus 1992, p. xv-xvi), but his later positive assessment of Freeman's brain dynamics (see Dreyfus 2007 and Freeman 1999) confirms at least his hope that GOFAI will be superseded by a more suitable approach to cognition, that is, by an approach closer to neuroscience, despite the endless efforts for a symbolic top-down approach.

This view also chimes with Dennett's unbridled enthusiasm for connectionist computing models, which provide the basis for the development of an understanding of cognition in fully accord with a bubble-up theory of creation (see Dennett 2009, p. 10061). Intelligence is a slippery concept and it must be admitted that GOFAI is the bulwark of a rationalistic tradition in the philosophy of mind which favors the existence of a language of thought or Mentalese in the style of Fodor (Fodor 1975; Fodor 2008). In accordance with such a view, thought in human brains is organized through a set of representations of environmental features, which can be coupled with certain aspects of language, that is, thought is altogether syntactic. So when it comes to GOFAI cognitive architectures it actually makes sense to pose the question as to how the computer can represent certain objects of the external world. This is, however, not the case with neural networks, whose framework based on states of activation,

output functions, patterns of connectivity, and activation and learning rules (see Rumelhart 1989, p. 209), contrasts sharply with GOFAI inspired architectures. In fact, connectionist models excel at precisely those phenomena (speech perception, visual pattern recognition, etc. See Smolensky 1989, p. 233), which proved unreachable for GOFAI. So it is not at all hard to see why Dennett has joined the enthusiastic bandwagon toward connectionism. In the first place, in connectionist models computation is performed at the level of simple units, which in turn do not require any representation. Representation is just the whole product ensuing from the network's entire activity. Connectionist models are then thought to bear a striking resemblance to the workings of the brain, which consequently would allow for the emergence of consciousness as a resulting product of multiple uncomprehending robots or neurons—a view Dennett is very fond of, because, if successful, it will handsomely concord with Darwin's strange inversion of reasoning and Turing's own computational inversion.

A History of First-Step Fallacies

It might be true, as Crane asserts, that it is "an empirical or scientific question whether our minds have a classical Mentalese-style architecture or some mixture of the two—or, indeed, whether our minds have any kind of computational structure at all." (2010, p. 167) On Dreyfus account, connectionism is promising because it drew its inspiration, not from rationalistic philosophy, like GOFAI, but from neuroscience. Even if critics are often prone to remark that connectionism is just another version of associationism, "they may have overlooked the capacities of the most sophisticated neural networks" (Dreyfus & Dreyfus 1988, p. xii).

And yet Aizawa (1992) has identified a serious weakness in Dreyfus's attempt at drawing a clear line between rationalistic inspired GOFAI and neuroscientific, cerebral, purported holistic, non top-down representational neural networks. Connectionism, says Aizawa, has also drawn a lot from this rationalistic philosophical tradition, which Dreyfus finds essentially misleading: be it from logical calculus, be it from atomism and logicism (see 1992, p. 304-305).[32] Connectionism has both a variegated history (see Medler 1998)—spanning at least three decades starting in the post-perceptrons

[32] Aizawa bases his case, as is to be expected, on the classical paper by McCulloch and Pitts (see 1943).

era[33] (see Minsky & Papert 1988; first ed. 1969)—and a copious literature.[34] According to Aizawa (1992, p. 304), Dreyfus's attempt to differentiate between GOFAI and connectionism on the basis of an overt rejection of rationalism is nothing more than a caricature. Wittgensteinian holism, which Aizawa ascribes to Dreyfus as his fundamental presupposition, is straightforwardly noncognitive. So construed, "state changes that do not involve changes in atomic or molecular representations are non-cognitive," which means that "a holistic theory of cognition is simply not a conceptual possibility." (1992, p. 309) Given that connectionists have been mostly concerned with modeling neuronal processes,[35] it comes as no surprise that both Dreyfus and his brother Stuart (Dreyfus & Dreyfus 1988) have argued that GOFAI is concerned with making a mind that represents the world creating a suitable model of it, whereas connectionism attempts to understand the brain. Notwithstanding Aizawa's idea that whatever holism is, it should not be counted as pertaining thematically to cognitive science, it is arguable that ruling something out as noncognitive constitutes the whole story. This shall be discussed in detail in later chapters, but the question should be raised: if these noncognitive processes are indeed fundamental for human intelligence, doesn't that at least suggest that cognitivism could be false? And if not false in its entirety, at least incomplete to a certain extent?

Be that as it may, the truth is that Dreyfus (see his introduction to 1992) has also identified a series of difficulties and problems in connectionism, which continue to haunt AI, even under the rubric of the 'connectionist revolution.' According to his interpretation, for all the theoretical acumen of neural networks researchers, "the commonsense-knowledge problem resurfaces in this work and threatens its progress just as it did work in GOFAI" (1992, p. xxxvi), and this because "neural networks are almost dependent upon human intelligence as are GOFAI systems" (1992, p. xxxix). So the problem of assigning relevance from the top-down, that is, from the designer's point of view, is lurking in wait for

---

[33] Old connectionism differs from new connectionism in that the former could not combine two types of networks: on the one hand, trainable networks, such as layer perceptrons, that are computationally limited and, on the other hand, large computationally hardwired powerful networks (see Medler 1998, p. 77). Connectionist ideas, nonetheless, can be traced back even to the 19th century. It is often noticed that ideas underlying connectionism held sway first on the basis of an associationist psychology combined with neurology, for example in Herbert Spencer's and Williams James's work bearing, by happenstance, the same title: *The Principles of Psychology*.

[34] A whole myriad of references is provided by Boden (see 2006, pp. 1453-1586).

[35] There is more to connectionism, of course: "There were those (like Rosenblatt and the PDP [parallel distributed processing] group) who leant towards psychology and/or neurobiology. And there were those (like Widrow and Hopfield) who came from engineering, physics, computing, and/or mathematics" (Boden 2006, p. 959-960).

connectionism when it comes to generalization. For Dreyfus, the problem is as recalcitrant as ever for an intelligent network, because "for a given classification task, given sufficient examples of inputs associated with one particular output, it should associate further inputs of the same type with that same output. But what counts as the same?" (1992, p. xxxvi) The idea behind this criticism is that whatever counts as a specific definition of 'type' required for a reasonable generalization, it is no other than the network's designer who must advance a specification of this sort. If, however, the network produces an unexpected association, the network's designer is ultimately responsible for specifying what counts as contextual and what not. Dreyfus's factual example for this (see 1992, p. xxxvi) is a neural network trained by the military to recognize tanks in a forest, which, at first, presented impressive results. The network could not only successfully discriminate, out of pictures that were shown to it, between a forest with tanks and one without them, but it could also generalize its knowledge to pictures that had not been part of the training. Despite the impressive prowess of the network, the sense of success was rather ephemeral, because the network was constantly failing to discriminate between new pictures of trees with tanks behind them and the new pictures of just plain trees: "the mystery was finally solved when someone noticed that the original pictures of the forest without tanks were taken on a cloudy day and those with tanks were taken on a sunny day. The net had apparently learned to recognize and generalize the difference between a forest with and without shadows!" (Dreyfus 1992, p. xxxvi). This illustrates clearly what Dreyfus argues about commonsense or the sense of relevance of a situation: "a network must share our commonsense understanding of the world if it is to share our sense of appropriate generalization." (*idem*). Therefore, it is inaccurate to regard Dreyfus's enthusiasm for connectionism as more noticeable than it actually is. As a matter of fact, in conformity with Dreyfus's assessment, "it looks likely that the neglected and then revived connectionist approach is merely getting its deserved chance to fail" (1992, p. xxxviii).

Dreyfus's depiction of the predicament in which AI finds itself as a research program has been —and continues to be—radically negative: the talk is always of a waning of the program, an impasse, a degeneration. This is due to his unshakable belief that AI's history can be best described resorting to Bar-Hillel's idea of the 'first-step fallacy': "first-step thinking has the idea of a successful last step built in. Limited early success, however, is not a valid basis for predicting the ultimate success of one's

project. Climbing a hill should not give one any assurance that if he keeps going he will reach the sky" (Dreyfus 2012, p. 87).[36] Nevertheless, after all the theoretical shortcomings which have come about and are still problematic for building true AI systems, the iron will of AI researchers towards flamboyant predictions has not ceased and not an iota of what Dreyfus calls the unfounded optimism of AI has changed.

This groundless optimism, expecting the best out of not so much,[37] is best exemplified in today's speculations about the singularity: the event, due to the exponential growth of computing power, which will inexorably lead to machines becoming self aware with unpredictable consequences. This is Kurzweil's gospel: the singularity refers to the *Age of Spiritual Machines* or *When Computers Exceed Human Intelligence* (see 1999). According to Kurzweil's own predictions, by the year 2019 one should expect, for example, automated driving systems being installed in most roads, or people beginning to have relationships with automated personalities and using them as companions, teachers, caretakers, and lovers! (see 1999, p. 279).[38] In addition, there will be "widespread reports of computers passing the Turing test" (*idem*). Ten years later, by 2029, a computer will have the capacity of one thousand human brains, there will be direct neural pathways for high-bandwidth connection to the human brain, and there will be growing discussion about the legal rights of computers and what constitutes being human. Machines, finally, will claim to be conscious and these claims will be largely accepted (*idem*). This will be nothing less than what Chorost (see 2011) has called *World Wide Mind* or the coming integration of humanity, machines, and the internet. And if these ideas seem far-fetched, how about conceiving of computers modeling human consciousness, which "will permit us to download our personalities into non-biological substrates. When we cross this... bridge, we'll become information. And then, as long as we maintain copies of ourselves to protect against a system crash, we won't die" (Wolf 2008. Quoted

[36] Dreyfus's brother, Stuart, quipped: "It was like claiming that the first monkey that climbed a tree was making progress towards landing on the moon" (Dreyfus 2012, p. 92).

[37] Hilary Putnam comes to mind, when he—unimpressed—deemed AI resorting to one of Shakespeare's comedy titles: much ado about not very much (see 1988).

[38] As widely reported by the press, in 2016 a Tesla Motors' autopilot car crash continues to dissipate any hopes for near future autopilot features being massively incorporated into everyday life in vehicle driving.

by Dreyfus 2012, p. 88). For Dreyfus this is not only a little unlikely,[39] but, if one reflects upon the history of AI, it certainly constitutes a blatant example of first step thinking.

The history of the extravagant predictions made by AI researchers has been only sketched in this chapter and the reference to GOFAI systems might give the impression that this is, after all, old news. But even today, a philosopher with a prominent academic post, like Nick Bostrom (Oxford professor and director of The Future of Humanity Institute), has it that it is important to reflect upon an AI takeover scenario: "A machine superintelligence might itself be an extremely powerful agent, one that could successfully assert itself against the project that brought it into existence as well as against the rest of the world." (2014, p. 95) Therefore, Dreyfus's (2012) new take on this history as a progression of first-step fallacies, might be of use. According to Dreyfus, the history of AI research might be recounted as a series of first-step fallacies.

· *First-step fallacy one* comprises basically the beginnings, that is, the new field founded by Newell and Simon of cognitive simulation which put forward for the first time the physical symbol system hypothesis (2012, pp. 89-90).

· *First-step fallacy two* came to the fore when a series of attempts were made in order to overcome the frame problem, and so programmers working under Minsky were painstakingly trying to fill a huge database with all facts about the world that could be imagined as forming part of every relevant situation (2012, pp. 91-92).

· *First-step fallacy three* is best exemplified by the emergence of experts systems, when AI researchers redefined themselves as 'knowledge engineers' and devoted their efforts to building systems—again, resorting to huge databases and brute computational force—divorced from everyday life (2012, pp. 92-93).

· *First-step fallacy four* was the purported big leap taken towards programming Cog, a humanoid robot that nowadays is exhibited in a museum (2012, pp. 93-94).

· *First-step fallacy five* is represented by Lenat's efforts to tackle the commonsense knowledge problem by means of the CYC (from encyclopedia) project: an immense knowledge base containing tens of

[39] Dreyfus actually makes fun of this when he quotes a character of McLeod's novel *The Cassini Division* (1988), who dismissively refers to the singularity as "the Rapture for nerds" (Dreyfus 2012, p. 99).

millions of entries, attempting thus to assemble a comprehensive ontology of everyday commonsense knowledge, which would purportedly enable AI applications to perform human-like reasoning (Lenat, Prakash & Shepherd 1986). When Dreyfus engaged in a debate on CYC in 1992 in the introduction of the MIT edition of his major work, Lenat still maintained his hopes up that he could somewhat finally exorcize the demons of one of the most fundamental problems haunting AI research. But as Dreyfus projected, CYC was still indebted to GOFAI and its assumptions.

· Finally, *first-step fallacy five* is very daring, for it leaps ahead to futuristic fantasies even without countenancing the current poor state of machine intelligence in comparison to human cognition. What is most striking about this idea is perhaps not Kurzweil's extravagant claims and predictions about the singularity, but that philosophers have taken them seriously.[40] Chalmers (2010)—the first philosopher that comes to mind in this regard—thinks that the singularity should be of great interest to academic philosophers, cognitive scientists, and AI researchers alike. The fact that this is not the case, is for him a matter of regret. Now, is it a real possibility? According to Chalmers, the progression that goes from AI (human level) to AI+ (greater than human level) and, finally, to AI++ (super intelligence), should not be easily dismissed:

> Given the way that computer technology always advances, it is natural enough to think that once there is AI, AI+ will be just around the corner. And the argument for the intelligence explosion suggests a rapid step from AI+ to AI++ soon after that. I think it would not be unreasonable to suggest 'within years' here (and some would suggest 'within days' or even sooner for the second step), but as before 'within decades' is conservative while still being interesting. (2010, p. 13)

Lest that not be enough, Chalmers goes on to offer the following logical argument in three premises:

> [1] There will be AI+.
>
> [2] If there is AI+, there will be AI++.
>
> [3] Therefore, unavoidably there will be AI++.

But as Dreyfus notices, "premise 2 is an explicit example of the first-step fallacy" (2012, p. 96). Furthermore, "it does not occur to Chalmers that his argument requires a step *before* his first step: that there be a successful AI program" (2012, p. 97), which there isn't. AI is simply not around the corner. In conclusion, "there is, in fact, *no reason to think that we are making progress towards AI or, indeed, that AI is*

[40] In a recent interview with Tim Crane, Huw Price, Bertrand Russell Professor of Philosophy at Cambridge, has claimed: "I do think there's a strong case that some time over the next century or two we are going to encounter a major transition when we do really have AI which is potentially, in some senses, a lot smarter than we are" (Price 2012).

even possible, in which case claiming incremental progress towards it would make no sense" (*idem*). This glib notion of AI making progress towards accomplishing human-level intelligence strikes one as being deceitful, although it was instrumental decades ago when it was all about getting funding for research (see Minsky and Papert 1972). But Dreyfus is right in that there is no reason to assume that current achievements constitute a step in the right direction on a continuum that will lead to AI, let alone to AI+ or AI++.

If progress towards accomplishing machine intelligence resembling human-like cognition has been so far negligible, one had better map out those problems that are still open and in need of theoretical clarification. Floridi (ed. 2008) has recently presented a series of researchers with the question as to the most important open problems concerning computational and informational issues (another way of designating 'philosophy of AI'). Among them, the frame problem, thinks Dennett, still looms large (p. 59). McCarthy is still puzzled by the representation of commonsense information and reasoning (p. 110). Boden suspects that a fundamental conceptual revolution, within both neuroscience and the philosophy of mind, may be needed before we can understand qualia (p. 9). Winograd deems the relationship between computation and the informational activities of the human brain/mind highly problematic (p. 175), and a big etcetera. If AI has contributed to philosophy, it has done so by providing a panoply of problems, rather than solutions. This is why it remains to be asked if one should see the prospects of AI research as merely forlorn, plagued by devilish difficulties and misbegotten ideas. In this vein, one could pose a question and give an answer, and this by echoing Putnam: "Has AI taught us anything of importance about the mind? I am inclined to think the answer is no. I am also inclined to wonder, What is all the fuss about?" (1988, p. 267). Some philosophers, like Dennett, see the strength of AI approaches precisely in this panoply of failures, because at last philosophy has become experimental. Perhaps it is wrong to conceive of the mind essentially as representational, or what not, "but one simply cannot tell without actually building the models and testing them" (Dennett 1988, p. 290). After all, according to Dennett, on what in-principle is possible or impossible, "the philosopher's answer is always, 'More a priori analysis and argument'. The AI researchers' answer is, Build it and see" (1988, p. 291).

Despite ubiquitous accusations as regards Dreyfus's alleged jaundiced derision and obstructive attitude towards AI, his contribution to the debate offers also hints of the way one could tackle those

problems with which philosophers have grappled for generations and which AI researchers have been trying to resolve by technical means. It is not accurate to suggest that Dreyfus has been only toying with in-principle or dogmatic arguments, making a windfall from another's misfortune, namely that of AI research. So much, then, for criticism. Has Dreyfus anything more constructive to offer? Dreyfus's interpretation of skillful coping phenomenology is tantamount to his mayor contribution to this discussion, and that shall be the topic of the next chapters.

| 4 |

THE PRIMACY OF PHENOMENOLOGY OVER LOGICAL ANALYSIS

The Reasons of the Heart

By the mid 1980s, when Dreyfus authored a book along with his brother Stuart on the power of human intuition and expertise in the era of the computer (as the book's subtitle goes, see Dreyfus & Dreyfus 1986), he had come to the conclusion that one of the aspects that had stymied progress in machine intelligence was, once again, an unchampioned philosophical assumption: that which grants primacy to 'knowing that' over 'knowing how.' On Dreyfus's view,[41] this overall conception was most pervasive in AI and should be corrected precisely in the heyday of expert systems, which were at that time part of a knowledge engineering enterprise aiming at building knowledge-based systems and programs (Feigenbaum 1977).

On this account, "the key idea of expert systems is that of making the knowledge that underlies expertise *explicit*" (Fox 1996, p. 180). Starting from the mid 1960s through the 1970s and 1980s, hopes were at their highest level as programs were beginning to be developed that could purportedly simulate specialist knowledge gleaned from human experts, of which DENTRAL (an expert system for determining complex organic molecules) and MYCIN (an expert system for the diagnosis of infectious blood diseases) were pioneering examples.[42] The foray of AI practitioners into knowledge engineering was supposed to assist the project of gradually getting rid of the always onerous services of real, flesh and blood experts, whose knowledge and expertise implemented in computer programs would open the prospect of computer-aided decisions based on an enormous knowledge database that no real

[41] Although *Mind Over Machine* (1986) was written jointly by Hubert and Stuart Dreyfus, quotes from it shall be (for stylistic reasons) presented from Hubert Dreyfus's point of view. This strategy has also been followed by the Dreyfus brothers throughout the book.

[42] There are, of course, many more so-called expert systems. In 'Expert Systems Versus Intuitive Expertise' (chapter 4 of Dreyfus & Dreyfus 1986, pp. 101-121), some of them are critically put into question: for example, COGEN, R1, MACSYMA, PROSPECTOR, INTERNIST-I, PUFF and RECONSIDER. Dreyfus quotes Schank's assessment of expert systems in order to make it his own: "the words 'expert system' are loaded with a great deal more implied intelligence than is warranted by their actual level of sophistication" (Dreyfus & Dreyfus 1986, p. 101). See also Fox's (1996) summary of the first three decades of expert systems.

person could manage on her own. As Simon pointed out to Crevier reflecting on the reasons for the belated development of knowledge engineering, in the first decades of AI, researchers "did steer away from problems where knowledge was the essential issue" (Crevier 1993, p. 147), because, simply put, no large database could be built with the computers which were then available. So AI research, admits Simon, had to focus for a time on mere toy tasks: "Quite deliberately we did a lot of our work in the early days... on toy tasks" (Crevier 1993, p. 146). Feigenbaum, who participated in the development of DENTRAL, argued that AI researchers should get real, meaning they should better get real tasks to work on: "You people are working on toy problems. Chess and logic are toy problems. If you solve them, you'll have solved toy problems. And that's all you'll have done. Get out into the real world and solve real-world problems" (Feigenbaum & McCorduck 1983, p. 62). 'Getting real', in this sense, amounted to designing intelligent systems that could actually be of use in practical matters. It is no wonder then that both the military and the industry were very interested in the prospects for the foreseeable future of these new knowledge-based systems. For AI researchers working now on expert systems it was now all about searching for ways to make knowledge available to their computers in larger and larger amounts.

But if expert systems were to prove germane—rightly deserving the name of 'knowledge engineering,' one might add—human expertise, whether in medical diagnoses or in military intervention, should little by little become more or less dispensable. The fact, however, that nowadays we do not seem to have gotten rid of the dexterity provided by physicians, nurses, and military commanders, speaks for the need of human expertise, in fact, of the implacable irreplaceability of human prowess. Boden (2006) has provided an example of why trusting programs instead of real people may have resulted in an ominous warfare disaster, when a nuclear red alert in the USA during the Cold War was caused by an unknown object on the horizon:

> The reason why this frightening episode didn't escalate was that someone ruminated that the Soviets hadn't been making especially threatening remarks recently. The norms of political behaviour even during the Cold War, therefore made it highly unlikely that this mysterious object was a Soviet attack. And the same rules deemed it inadmissible to launch defensive nuclear weapons on the basis of such weak—i.e., *politically* implausible—evidence. Accordingly, the computer was overridden. (The unknown object eventually turned out to be the rising moon.) (2006, p. 1019)

This anecdotal example makes clear how a shred of common sense, which expert systems overtly lack, can be of utmost importance when it comes to expert know-how being required in critical situations. The problem is, as Boden has put it, "whether professional expertise is describable, even in *words*, in its entirety" (2006, p. 1016). The large and even impressive range of explicit knowledge with which a knowledge-based program is provided (with all its heuristic, problem-solving rules of thumb) comes short to actual human expertise when the slightest sense of being in a human situation is lacking. Seemingly, the commonsense knowledge problem is here as recalcitrant as ever, because expertise and professional knowledge are apparently not bound up with explicit facts and theories. Rather, much of the expert's knowledge consists "in informal heuristics developed over the years, rarely verbalized and almost never communicated" (Boden 2006, p. 794).

Now, on Dreyfus's view, it is all the more understandable that, as part of any scientific enterprise, wrongheaded hypotheses be discarded, being superseded later on by better and more accurate ones. But "unfortunately, what has always distinguished AI research from a science is its failure to face up to, and learn from, its failures" (2007, p. 249). Moreover, many decades of AI research have "lived up to very few of its promises and [have] failed to yield any evidence that it ever will. The time has come to ask what has gone wrong and what we can reasonably expect from computer intelligence" (Dreyfus & Dreyfus 1986, p. xvii). After all, knowledge engineering poses the question: "What [are] the heuristics of performance in hypothesis formation?" (Feigenbaum 1992, p. 7). Knowledge engineers were accordingly entrusted with the task of making sure that the computer has all the knowledge needed to solve a problem. But how can this requirement be met? Is expert knowledge a cluster of millions of data that can be made explicit to deliberative rationality? In fact, can this knowledge be made explicit in order to feed up a program with its verbalizations? Dreyfus is of the opinion that the formulation of every possible rule underlying performance in a given task domain would still leave out other factors, which must be taken into account if one is to understand practice. These factors constitute the background which is always presupposed in human skillful activities and are denizens in the fringes of consciousness (Dreyfus 1992, p. 103). In order to make this point, Dreyfus draws from Polanyi's (1962) assumption that there is a great deal of knowledge presupposed by science that is not—and cannot be—made explicit. According to Polanyi's view of scientific

discovery, unlike what objectivists would let us believe, the grounds on which science is pursued "is determined at every stage by indefinable powers of thought" (1966, p. 1). This accounts for the fact that "tacit knowing is the fundamental power of the mind which creates explicit knowing, lends meaning to it and controls its uses" (1966, p. 18). Moreover, "any attempt to gain complete control of thought by explicit rules is self-contradictory, systematically misleading and culturally destructive. The pursuit of formalisation will find its true place in a tacit framework" (*idem*).

Hence Dreyfus's motto (see the Prologue to Dreyfus & Dreyfus 1986) taken from Pascal introducing his account to deal with that which in AI appears as unimportant: "The heart has its reasons that the reason does not know." The clarification of these reasons which most of the time are not made explicit—simply because doing it would amount to interrupting what is at stake in performance—begins, on the one hand, with understanding that 'knowing how' cannot be reduced to 'knowing that.' On the other hand, one should not be utterly oblivious of the aforementioned interruption of absorbed coping, because when thematization arrives on the scene, performance itself has been most probably interrupted from its natural flow. As a result, Dreyfus's strategy consists in analyzing knowledge acquisition: what he calls "our phenomenology of knowledge acquisition" (Dreyfus 1987, p. 30). Every expert has begun being a clumsy novice. How is it then that she can become an expert worthy of the name? Is it true that, in spite of appearances, "the mind and brain must be reasoning—making millions of rapid and accurate inferences like a computer [?]" (*idem*). The main assumption behind the idea of expert systems, that experts must be making inferences from stored data, must be questioned, admitting the role of involvement and intuition in the acquisition and application of skills. On Dreyfus's view, there is "no reason to cling to heuristic program as a model of human intellectual operations" (1987, p. 31).

In the light of the above-mentioned objections raised by Dreyfus, five stages are to be considered in order to understand knowledge acquisition.

Stage 1: A novice learns at first to manipulate unambiguously defined context-free elements by precise rules, a procedure that Dreyfus christens 'information processing' (Dreyfus & Dreyfus 1986, p. 21). There is no situational awareness at this stage and therefore rules ignore context outright. Novices

usually cling strictly to the handbook of rules they have been given, which thwarts most of the time their being flexible to changes.

Stage 2: An advanced beginner must now cope with real situations, which improves performance to an acceptable level. This stage provides the learner with practical experience which cannot be taught by sheer rule-following and context-free facts. This is the stage in which features that at first were not made explicit by the rules, begin to be recognized. So mere handbook rules begin to appear embedded in a situation, which requires more than the sheer application of them.

Stage 3: Someone exhibiting competence is able to deal with a number of both recognizable context-free and situational elements in a real world circumstance (see Dreyfus & Dreyfus 1986, p. 23). And yet, the competent performer keeps seeing the situation as a set of facts that can be grappled with the adoption of a hierarchical procedure of decision-making: "A competent driver, for example, is no longer merely following rules designed to enable him to operate his vehicle safely and courteously but drives with a goal in mind" (Dreyfus & Dreyfus 1986, p. 24). Dreyfus's idea here is that proponents of problem-solving strategies (generally cognitive scientists, psychologists and AI researchers) can only by these means characterize competence but not yet performance. On this view, all intelligent behavior tends to be characterized under the rubric of problem-solving.

Stage 4: Proficiency implies that the learner of a new skill has managed to cope with situational relevance and has consequently made conscious choices in the form of reflection upon several alternatives. She has now perspective, which allows for certain features to appear as salient as a consequence of past interactions with similar situations. This *personal* perspective is most likely acquired when coping several times with the same situation. Past interaction with a situation induces the proficient performer to form for herself an assessment of what can and cannot, and what should and should not be done in certain contexts.

Stage 5: Finally, "an expert generally knows what to do based on mature and practiced understanding" (Dreyfus & Dreyfus 1986, p. 30). Thus experts have at their disposal the ability to discriminate relevance and nonrelevance among a myriad of situations; a skill that, incidentally, resists explicit thematization: "we doubtless can discriminate many... situations than we have words in our

vocabularies. Consequently, such grouped situations bear no names and, in fact, seem to defy complete verbal description. With expertise comes fluid performance" (Dreyfus & Dreyfus 1986, p. 32).[43]

Given that experts exhibit a level of skill characterized by a fluid, involved (and absorbed, nonthematic) kind of behavior, this would cast into doubt the idea that the detached, context-free, problem-solving, and information-processing strategies customarily proposed by computer scientists can somehow be transformed into meaningful, situational coping. There is no reason to suppose that this is the correct approach to skill acquisition and, in fact, the recognition of this misleading approach may well render false the theoretical assumptions underlying expert systems implementation. The stages of skill acquisition put into question the explanatory powers of representationalism: "if the skill story I just told is correct, however, the problem of association of representations of an object can be avoided. What one has learned appears in the way the world shows up; it is not represented in the mind and added on to the present experience" (Dreyfus 2002a, p. 373). Drawing conclusions from Merleau-Ponty on this regard, Dreyfus insists that "what the learner acquires through experience is not *represented* in the mind at all but is *presented* to the learner as a more and more finely discriminated situation, which then solicits a more and more refined response" (*idem*).

This is why some philosophical lessons are here to be learned. At stake is how to understand those 'reasons of the heart' that deliberative reason—and its *a posteriori* verbalizations—ignores. However, Dreyfus's resort to the nonrepresentational stance provided by intuition as something which defies analysis is wont to be seen as magic by scientists and science-oriented philosophers alike. What is meant by intuition in this regard is not to be understood as the mere misnomer referring to a process which is badly understood. On the contrary: "*intuition or know-how, as we understand it, is neither wild guessing nor supernatural inspiration, but the sort of ability we all use all the time as we go about our everyday tasks*" (Dreyfus & Dreyfus 1986, p. 29). The lack of experience that one is to find in novice performance is due to the attempt to decompose patterns—which are immediately not recognized as such—into component features. Those patterns are at first provided as analysis of the situation, not as the involvement in a situation. The converse is true of experts, who have the ability of holistic discrimination and association (Dreyfus & Dreyfus 1986, p. 28). Therefore, holistic template matching and association

[43] On the five stages of skill acquisition, see also Dreyfus (2002a), pp. 368-372.

based on past experience as defining characteristic of expertise, cannot be derived from information processing and mere rule-following. Drawing heavily from a phenomenological tradition that privileges perception and bodily prowess over merely cognitive—namely, 'mental'—activities,[44] Dreyfus prefers skillful embodied coping as exemplary cases over machine-intelligent, information-processing ones:

> A boxer seems to begin an attack, not by combining by rule various facts about his body position and that of his opponent, but when the whole visual scene in front of him and sensations within him trigger behavior which was successful in an earlier similar situation. We call the ability to intuitively respond to patterns without decomposing them into component features 'holistic discrimination and association.' (Dreyfus & Dreyfus 1986, p. 28)

It should be obvious now that calculative rational procedures, which are all the time proposed by AI researchers as fundamental to intelligent behavior, lead one to regress to novice performance or to competent performance at most, but *never*—and this is the gist of Dreyfus's argument—to expertise. In novice performance, a sort of monitoring sticking strictly to the handbook of instructions and rules is constantly present in order to constantly observe which actions are in need of reinforcement or correction. This means, that "a portion of the mind is thus responsible for the fine tuning or disaggregation of discriminable classes for more effective guidance of future behavior" (Dreyfus & Dreyfus 1986, p. 40). Even if this species of monitoring can be traced back to the learning process, the presence of it is not at all times extant, for "there are rare moments, however, when all monitoring ceases. We are referring to those brief periods of what is sometimes called 'flow,' when performance, accompanied by a feeling of euphoria, reaches its peak" (*idem*). Typically, this flow is to be found in the experience of performers, for instance in an sportsperson's sense of what needs to be done in order to accomplish a certain aim, say a footballer's attempt to score a goal. So the 'flow' is not "a sixth stage of mental activities that produce skilled behavior but rather the cessation of the monitoring activity that normally accompanies the higher levels" (*idem*). This means that 'knowing that' is just *a posteriori* reflection upon 'knowing how'; the latter being fundamental and basic, not conversely.

Dreyfus is hence conscious that his thrust against deliberative rationality renders his account of skillful human activity away from the mainstream philosophical tradition. According to him, there is a transformation of ancient *logos* into *ratio* and, consequently, into 'reckoning' which we are accustomed

[44] Some classical phenomenological examples come to mind. Among the most salient ones are Heidegger's dealings with equipment or *Zeugzusammenhang* (see, for example, the always cited 'hammering with a hammer,' SZ § 15, p. 69) and Merleau-Ponty's many exemplary cases of bodily comportment (see Merleau-Ponty 2005).

to taking for granted as being the kernel of human rationality (Dreyfus & Dreyfus 1986, p. 205). But Dreyfus is not pleading for a sort of feeling-romanticism that rejects all rationality, but just the kind of claim that reduces rationality to technological explanation: "the question is whether we are going to accept the view of man as an information processing device or whether we are still enough in touch with our pre-Platonic essence to realize the limits of the computer metaphor" (*idem*). Nonetheless, deliberative rationality is not to be rejected *tout court*: "put in its proper place rational deliberation sharpens intuition" (*idem*). So the study of knowledge acquisition pursued by Dreyfus reveals that the world is not a collection of objects, which are captured in relation to a thinker. Dreyfus seems thus to agree *verbatim* with Merleau-Ponty regarding perception:

> We cannot apply the classical distinction of form and matter to perception, nor can we conceive the perceiving subject as a consciousness which 'interprets,' 'deciphers,' or 'orders' a sensible matter whose ideal law it would possess. Matter is 'pregnant' with its form, which is to say that in the final analysis every perception takes place within a certain horizon and ultimately in the 'world,' that both are present to us practically rather than being explicitly known or posited by us... (2007, p. 89)

For Dreyfus, what expert systems assume is precisely the converse: that the perceiving subject deciphers the objective world by means of information retrieval from millions of data, the world being thus in turn nothing but a collection of objects and facts, as though it were natural to superimpose, as Merleau-Ponty would have it, "a world of ideas on the perceived world" (2007, p. 89). But as the analysis of knowledge acquisition shows, conscious thought about one's own activities rather leads to degradation of performance: "here you fell victim of 'knowing that' as it interrupted and replaced your 'knowing how'" (Dreyfus & Dreyfus 1986, p. 17).[45] Know-how consists in forgetting how one actually does it, and this is why nonthematization of the current flow of the ongoing situation is fundamental for satisfactorily coping with it, that is to say, for exhibiting robust prowess and expertise, which appears wonderfully ordinary and easy in human agents. This seems to bring Dreyfus's musings in close connection to the proclamation of practice as more fundamental than theory, but he is looking for more than that. He is looking rather for the depths and basis of propositional thinking.

[45] The reader might have experienced forgetting the PIN code of a bank account, precisely when trying to remember it *propositionally*, as though the flow of that sort of knowing were interrupted by thinking about it! When that happens to me, I force myself not to think about it and let the body do the work.

Absorbed Coping

Before laying out the details concerning absorbed coping, a caveat is in order. As a sort of 'Heideggerian', Dreyfus is known for his animosity towards Husserl, which he might have contracted from Heidegger's own critical attitude towards the father of phenomenology, but most certainly from an Anglo-American construal of Husserl (see Føllesdal 1979) as a philosopher who favors an overemphasis of detached contemplation over situational coping. Surprisingly, this line of interpretation can even find in Husserl's philosophy a proto-Fodorian theory of mental representation.[46] McIntyre (1986) has argued, for instance, that Husserl's noematic *Sinn* is tantamount to Fodor's mental representations and that his so-called phenomenological reduction shares fundamental similarities with Fodor's methodological solipsism (p. 101); the latter being a doctrine that Dreyfus has castigated in what he calls Dasein's revenge, meant to "undercut the Cartesian prejudice that man is a subject embedded in the physical world" (1980, p. 78). Rather, "the pragmatic activity of taking-to-refer and claiming-to-be-true takes place against a background of already entrenched social practices" (*idem*).

Dreyfus also falls prey to this barely phenomenological Husserl interpretation and hence is not really able to construe classical phenomenology as Heidegger did—as the very "possibility of thinking" (GA 14, p. 101), albeit in need of a radicalization to salvage it from its commitments to the metaphysical tradition. Actually, this Husserl interpretation has been ubiquitous in the Anglo-American reception of phenomenology. This is the case of the West Coast interpretation of Husserlian phenomenology, which has defended a Fregean interpretation of Husserl's theory of intentionality and conceives of the noema as an intermediary ideal entity (see Zahavi 2003, p. 58 ff). Varela, Thompson and Rosch, in a work (1993) whose partial purpose was at least to consider phenomenology as a fruitful philosophical companion to cognitive science, concluded that the Husserlian project was a failure (1993, p. 19) on the grounds that Husserl was a methodological solipsist (1993, p. 16), whose philosophy ignored embodiment (1993, p. 17) given that it was basically a very abstract representational theory of mind (1993, p. 68). They even went so far as to deem the Husserlian theory of the life-world

[46] See McIntyre (1986). See also Livingston (2005) on the "substantial historical and conceptual continuity between functionalism and phenomenology" (p. 31).

reductionistic and representational (1993, p. 117), and opted out of phenomenology, what ultimately motivated a turn to the Buddhist tradition of mindfulness-awareness meditation. However, Thompson has recently changed his mind and no longer holds that Husserl's project is a failure (2007, p. 414). Apart from Heidegger's largely uncharitable reading of Husserl, which explains the rising number of English speaking Heideggerians who openly launch venomous attacks on Husserl's phenomenology, Thompson credits Dreyfus's interpretation (1982) as the received view of phenomenology in America, which played an important role in informing his (and Varela's) misconstrued previous understanding of Husserl's phenomenology. On this view, Husserl would hold that even practical activity is object-oriented (Dreyfus 1982, p. 9) and he would conceive of intentional experiences as belonging to a special realm of representational entities (Dreyfus 1982, p. 1).[47]

Smacking of this interpretation of Husserl's philosophy as an antecedent to cognitivism, Dreyfus consequently conjoins Husserl's and Searle's accounts of intentionality as if both philosophical undertakings were made of the same stuff: that is, as if both philosophers were staunch proponents of mental representation. Accordingly, Dreyfus can accuse both philosophers of capriciously subjectivizing intentionality, which has a great many serious disadvantages to be borne carefully in mind. Indeed, Heidegger has noted that "the idea of a subject which has intentional experiences... encapsulated within itself is an absurdity which misconstrues the basic ontological structure of the being that we ourselves are" (GA 24, p. 89). Heidegger's idea in this lecture is that the traditional split between mind and world is artificial and a mere theoretical postulation. Conversely, "I cannot and must not ask how the inner intentional experience arrives at an outside" (GA 24, p. 89). On this account, Dreyfus's argumentative strategy consists in pointing out why Searle's formulation of the way the mind/world split is supposedly built into the experience of action must be questioned.

In *Speech Acts* (originally published in 1969), Searle argues that the essence of language lies not in its being only propositional content but fundamentally by consisting in performative or—drawing from Austin's terminology—illocutionary acts. Speech acts must therefore be conceived as intentionally acting according to rules, which Searle calls 'constitutive': "the different speech act types can then be

[47] As Thompson has noted (2007, p. 415), this misleading Husserl interpretation has been challenged by a number of philosophers. See Marbach (1993), Welton (2000) and Zahavi (2003; 2004). For other misuses of phenomenology in connection with Varela's neurophenomenology, see Lembeck (2010) and Ebinger (2012).

seen as providing different institutional possibilities within the institution of human language, and explaining the structure of speech acts is a matter of laying bare the constitutive rules" (2001a, p. 174). Searle then goes on to argue that much of the harder work in speech act theory was directed at answering how we get from the physics of sounds to the semantics and pragmatics of speech acts: "how do we get from the acoustic blast that comes out of the speaker's mouth to the illocutionary acts?" (*idem*). So the different possibilities of illocutionary (that is, performative) speech acts must be mapped out as one would proceed with a territory: by classifying the significant dimensions of differences between these acts (Searle 1976).[48] But given an account of every possible performative linguistic act implies the extensive resort to mental notions such as belief, desire, and intention. Searle does well when he admits that his theory of intentionality (see 2008a) owes nothing to Husserl's phenomenology, but instead came straight from his speech act theory under the influence of Frege, Strawson, Wittgenstein, and Austin (2001a, p. 175). This recognition is important because it is clear that, when referring to intentionality, Searle thinks about *intentions* much in the same line of Anscombe's project of making plain the character of human action and will (see 2000).[49] Searle's distinction between propositional content and illocutionary force, provides one with the possibility of carrying over to the structure of intentional states, for one can both assert that p, or ask whether p, or make a promise that p, just as one can also believe that p, wish that p, fear that p, etc. (2001a, p. 175). At the same time, the Searlean notion of intentional causation makes possible an analysis of the structure of willful or deliberative acts (as it would be perhaps better to characterize Searlean 'intentionality') in terms of the conditions of satisfaction, the direction of fit, as well as some other notions that Searle introduces in other works (see 2008b).

On discussing Searle's work, Dreyfus's attention was drawn to how on actions, as conceived from Searle's viewpoint, an 'intention in action' (a continuing representation on behalf of the agent during the action itself) was arbitrarily superimposed. Notable is also Searle's insistence that the agent must experience the casual connection between the intention in action and the bodily movement

---

[48] Searle (1976) has classified illocutionary acts by means of a taxonomy, correcting thus the previous one provided by Austin in *How to Do Things with Words* (1962), which Searle deems defective for its lack of clear criteria.

[49] Anscombe's *Intention* (published originally in 1957) was rendered into Germans as *Absicht*, which very handsomely illustrates why 'intention' is not to be confused with anything related to phenomenological intentionality, which—one could argue—is precisely Dreyfus's mistake when associating Searle with Husserl's phenomenology.

continuously: "indeed, according to Searle, the experience of acting is just the experience of the bodily movement being caused by the intention in action" (Dreyfus 1993, p. 21). Dreyfus has it that "Searle attempts a unique integration of logical conditions and phenomenological description" (*idem*). Moreover,

> Searle incorporates a phenomenological analog of this analysis into his account of action by maintaining that the experience of an action must include a direct experience of the causal relation between the intention in action and the bodily motion. He argues that both the prior intention and the intention in action are casually self-referential. They both include in their conditions of satisfaction the requirement that the intention to bring about a goal cause the goal-directed action. Thus an action is a bodily movement experienced as caused by my intention to perform it. (Dreyfus 1993, pp. 21-22)

But both components that Searle ascribes to an action, the intentional component (for example, the visual experience of something which is perceived) and the conditions of satisfaction (the presence of features of what is seen), might be actually absent in real situational action. Of course, one has a visual experience and what is seen shows its own features that can be detected by the visual experience itself, but this way of putting things might just be an abstract way of reflecting upon an action and not the experience itself. So what is really an action? What elements should be borne in mind when considering what human agents undergo in the process of acting? What is agency defined as the capacity to act in a world? Human agents are in the first instance absorbed, coping with the situation in which they are currently involved. They are certainly not thinking deliberately about what is being pursued, nor representing, nor, to be sure, nothing of that sort. Thus Dreyfus casts into doubt the idea that an intention in action—the deliberate will to act upon something—is the cause of one's movement. Instead, acting is "the experience of a steady flow of skillful activity in response to one's sense of the environment" (1993, p. 24). Not even when things go wrong for a moment does an action resort to deliberative reflection. It "relieves the tension" (*idem*) of a deviation if, say, our bodily movement deviates from its course. This means that in the final analysis it must be recognized that, when coping skillfully with a situation, "activity is completely geared into the demands of the situation. One does not distinguish one's experience of acting from one's ongoing activity, and therefore one has no self-referential experience of oneself as causing that activity" (*idem*). Of course, this is not to decry deliberative reflection postulating its nonexistence or to negate that something as self-referring is

thinkable, but merely to object to its merits being attributed to a single false premise: that which grants primacy to representation and deliberative rationality in human agency.

In order to dig into the sense of human action and to lay out its nonrepresentational character (also referred to as 'nonintentionalistic,' see Dreyfus [1993], p. 24), Dreyfus draws phenomenological insights not only from Heidegger, but also from Gurwitsch and Merleau-Ponty. What occurs in any action is not imposed by agents, it is "rather prescribed by the situation and its own structure... We find ourselves in a situation and are interwoven with it, encompassed by it, indeed just 'absorbed' into it" (Gurwitsch 1979, p. 67).[50] By this quote from Gurwitsch, which Dreyfus reproduces, it is meant that "*the experience of acting* has a *world-to-mind* direction of causation also" (Dreyfus 2001, p. 25), which is clearly lacking in Searle's account of intentionality. In fact, "we experience the situation as drawing the action out of us" (*idem*). In this vein, Dreyfus agrees with Merleau-Ponty that human beings are "empty heads turned towards one single, self-evident world where everything takes place" (2005, p. 413). Dreyfus has it that the mixture of a first-person and a third-person—an internal and an external—account of perception and action is unstable and merely due to an unchampioned commonsense prejudice (1993, p. 26). Accordingly, "Searle starts from the first-person experience and builds the third-person casual account into the intentional content of the experience" (1993, p. 27). But "phenomenology rejects common sense in the name of the phenomena of everyday involved perception and action" (*idem*).

A wide and variegated range of situations, namely skillful habitual activities such as riding a bicycle, driving a car, and playing tennis, shows that Searle's depiction of the intentional content of acting in terms of a representation of the action's conditions of satisfaction is unnecessary and misleading. Thus Dreyfus has noted how much more time human agents spend in this immediate coping mode, when compared to the deliberative, purposeful, subject/object, theory-laden and theory-oriented mode of consideration, which is most of the time only secondary. As a matter of fact, actions can be purposive without the agent entertaining any kind of purpose at all (Dreyfus 1993, p. 28). This

[50] It must be noted again that Dreyfus understands Husserlian intentionality as detached observation of lived experience. Very early, indeed, he was keen to set transcendental phenomenology, conceived of as detached, objective reflection upon experience, against "the crucial role of human involvement" (1967, p. 19). An early criticism of this Dreyfusian misinterpretation of intentionality, was offered by Gurwitsch: "we ask whether involvement as experienced does not refer to consciousness experiencing it" (1974, p. 11).

can be defined as "the phenomenon of purposive action without a purpose" (Dreyfus 1993, p. 31), which has been more often than not totally ignored by the philosophical tradition of theory-oriented explicit deliberation. Therefore, very much in agreement with Heidegger and with the phenomenological tradition with its preference for nonobjective phenomena, Dreyfus is denouncing how the insistence on deliberative rationality ends up obfuscating the phenomenon of the world. But precisely, being-in-the-world as originary intentionality, "amounts to a non-thematic circumspective absorption in references or assignments constitutive for the readiness-to-hand of a totality of equipment" (Heidegger SZ § 16, p. 76). Being-in-the-world is a fundamental determination, not of objects, but of human existence or Dasein, and the way human agents are in the world is mostly an absorbed, concerned kind of coping with that which is required by the situation at stake. It is the world, as the orienting background, which makes coping with things possible, and it would be wrongheaded to imagine a gap between the agent's comportment towards what is being pursued and the world as disclosed. And this, because "self and world belong together in the single entity, Dasein. Self and world are not two entities, like subject and object... but self and world are the basic determination of Dasein itself in the unity of the structure of being-in-the-world" (Heidegger GA 24, p. 422).

Dreyfus thus inveighs against Searle's idea that whatever intentionality is, it must be circumscribed through reference to plain acts of the mind, that is, not to absorbed circumspection in the whole of activity. Against the backdrop of these objections, Dreyfus summarizes how skillful coping differs from mindless, mechanical behavior, since that is the impression that might surface when speaking of a non-self-referential experience of agency:

> · Skillful coping is a mode of awareness, but one in which the agent is not aware of himself as separate from the world (1993, p. 34). Such a sense of a subject being confronted by objects is an abstraction, a theoretical construct, and hence nothing which can be found in the experience of the phenomena.
>
> · Comportment is adaptable and copes with the situation in a variety of ways (*idem*), because one responds to things on the basis of past experience: "one's behavior manifests dispositions that have been shaped by a vast amount of previous dealings, so that in most cases when we exercise these dispositions everything works without interruption" (*idem*).
>
> · Finally, only if the going gets difficult we pay attention and so switch to a deliberate subject/object attitude (*idem*).

Again, this notion of the interruption of the flow of experience being the cause of the theoretical attitude is very Heideggerian for Dreyfus to entertain. Searle can indeed be chided for his emphasis on

explicit, transparent, and pervasive 'intention in action', but in observing even one's own activity, there is a monitoring attitude which is not present when everything goes well. So why is one to suppose that this monitoring observation is present originally in skillful coping? This monitoring way of knowing, which Dreyfus calls 'knowing that' when objecting to the approach to knowledge in expert systems, is not fundamental. As Heidegger has it, "if knowing is to be possible as a way of determining the nature of the present-at-hand by observing it, then there must first be a deficiency in our having-to-do with the world concernfully" (SZ § 13, p. 61). Surely, when the door is broken, my attitude towards it changes completely and I can now consider it from the perspective of mere thinghood. But that does not shore up the argument that things appear to a mind as sheer objects lacking meaning which must be somehow superimposed on them by mental acts. Thus Searle's question as to how we get from the acoustic blast that comes out of the speaker's mouth to the illocutionary acts (2001a, p. 174) is artificial, for it constitutes no real philosophical problem. It is instead just an assumption which arises from abstraction and theory alone.

The Dreyfus-Searle Debate

For some time, a fruitful exchange between Dreyfus and Searle seemed possible, specially because of Searle's insistence that "intentionality, in general, and meaning, in particular, always depend on a set of capacities that are not part of meaning or intentional content" (2001a, p. 176); a phenomenon that Searle calls 'the Background.' As defined in Searle's book about intentionality, the background of everyday practice is "a set of nonrepresentational mental capacities that enables all representing to take place" (2008a, p. 143). This striking similarity, at least superficially, explains why Dreyfus and Searle even gave seminars together on intentionality and read each other's papers (Dreyfus 1999, p. 3). However, profound differences began to surface when Dreyfus attempted to link Searlean philosophy with that of Husserl. Of course, Searle's use of the term 'intentionality' along with his emphasis on first-person experience, gave the impression that he was working very close to what Dreyfus thinks are phenomenological insights, even when he was not aware of it (this not being aware of his [Searle's] being a Husserlian is, of course, Dreyfus's idea). Indeed, Searle's critique of computational explanations of the mind (see Searle 1985) according to which "computation is not *discovered* in nature but is *assigned*"

(2001a, p. 178), is also shared by Dreyfus. However, a seemingly peaceful coexistence of phenomenology and logical analysis turned into a scathing debate, with Kelly, a former student of both philosophers at Berkeley, mediating between them (Kelly 2005).

Searle (2000) has subsequently defended himself from Dreyfus's attacks by pointing out, first, that he is not acquainted with the phenomenological tradition, and hence sees no point in Dreyfus's insistence that he must be some sort of Husserlian. Second, he finds his work has been grossly misinterpreted by Dreyfus, who would be, as it were, simply arguing against a straw man, not making any interesting points against the actual philosophical arguments to be found in his writings. And finally, he deems phenomenology as "the first step but only the first step in logical analysis" (2000, p. 72). Being the first step means for Searle that phenomenology, as exposed by Dreyfus, is concerned with the way phenomena appear for the first time to human agents, and that amounts to saying that phenomenology is superficial, since logical analysis concerns itself with a far deeper structure. Accordingly, Searle drives a wedge between his approach and that of Dreyfus:

> When I speak of 'representation,' 'conditions of satisfaction,' 'causal self-referenciality,' and 'intentions in action' he [Dreyfus] thinks I am talking about the phenomenology of agents. I am not. I am talking about the logical structure of intentional phenomena, and the logical structure does not typically lie on the surface, it is not typically discoverable by mere phenomenology. (2000, p. 75)

Searle objects to Dreyfus's approach that human agents without intentions for their actions are unrealistic:

> According to Dreyfus we are supposed to accept that when he wrote this passage, and presumably also when he rewrote, edited, and proofread it, he *had no mental states whatever*: *no "beliefs, desires, intentions, etc.*" Frankly, I find the idea out of the question. I believe that when Dreyfus wrote the passage, he did so intentionally, that is, he intended to write that very passage. Furthermore I think he wrote the passage in the "*belief*" that it was true and with a "*desire*" to say the things he said. Mental states like belief, desire, and intention are so "involved" in the production of this passage that if he had not had them he would not have written the passage at all. Worse yet, I believe that all of this skillful coping was conscious. (2000, p. 77-78)

So for Searle, "skillful coping is intentional behavior right down to the ground" (2000, p. 81). Moreover, Searle boasts of his inability to understand Dreyfus's beloved skillful-coping examples as a rebuttal to his philosophy and jokingly asserts that "except in a few really weird epileptic cases, all skillful coping requires consciousness" (2000, p. 82). He is not doing phenomenology, let alone of the Husserlian kind, for he is not interested in how things seem at a certain level of appearance, but he is rather analyzing logically the phenomena, not how they merely show themselves, but how they work and

constitute experience. In the same vein, "philosophy starts with the facts of physics, chemistry, biology, and neurology. There is no going behind these facts to try to find something more 'primordial'" (2000, p. 90). The problem is, as Searle has it, that neither Husserl nor Heidegger seem to have anything relevant to say about physics, chemistry, biology, and neurology. They seem merely to think it is important to theorize about how things appear to human agents, the difference between them being reducible to the triviality that one thinks intentionality is a subject/object relation between a transcendental subject and an intentional object, whereas the other doubts about the existence of this relation (*idem*). But in the final analysis, Searle thinks both philosophies of what phenomenologically appears to human agents are irrelevant "to getting an adequate theory of the logical structure of the intentionality of biological brains encased in biological bodies" (*idem*). Even worse, on Searle's view, Dreyfus's always 'presupposed background' does not object correctly to logical analysis, because it constitutes a fallacy: that of assuming that practices cannot be logically investigated because their practical background is always presupposed (see 2000, p. 92). But "just as we use the eye to study the eye, language to study language, the brain to study the brain, etc. so we can use the practices to study the practices, and indeed we can, as I do, use the Background to study the Background" (*idem*). Phenomenology is bankrupt because "it can only deal with how things seem to me here and now in the immediate present" (Searle 2001b, p. 282).

But Dreyfus is not convinced that these Searlean objections really capture the gist of his critique. As it happens, debates between philosophers belonging to different traditions of thought appear more often than not as a dialogue of the deaf, each side of the dispute talking past each other. The point is, however, which account of experience captures the way human agents gain knowledge and skill through habitual activity. Raging from knowledge engineering to Searle's mental capacities as trapped "in the head" (1991, p. 291), the traditional idea that somehow the mind assigns meaning to brute facts encountered in a world of objects still holds sway mightily. On Dreyfus's account, conversely, the capacities and skills which make up the background are bodily, rather than mental (2000, p. 325). For Dreyfus, the point is not, as Searle asserts, that he is not attempting to practice phenomenology, never mind of the Husserlian kind. Even if it is true that Dreyfus's account of Searle as Husserlian is misleading, Searle would be nonetheless engaging in bad phenomenology (Dreyfus

2000, p. 327). Let us not forget that, albeit Searle's naturalistic outlook of philosophy to which he subscribes (attempting thus to dismiss phenomenology for its anti-naturalism), he is not engaging in the well justified scientific question as to how the brain processes acoustic blasts coming out of people's mouths. He assumes, unjustifiably, that human agents somehow experience meaningless noises that can later become the experience of speech acts by mental meaning assignment, that is, he adopts the view that human agents first and foremost encounter meaningless facts needing a sort of interpretative supplement on behalf of the agent. Therefore, Searle is engaging in bad phenomenology, for he is assuming a meaningless stance of experience which never existed in the first place. Dreyfus then seems to be arguing alongside Heidegger that meaningfulness comes first ('das Bedeutsame ist das primäre,' GA 56/57, p. 73).[51]

Searle favors indeed a functional account of meaning, according to which meaning is not intrinsic to the physical stuff of the universe, but is instead "*assigned* from the *outside* by conscious observers and users" (1996 p. 14). Given that Searle's conditions of satisfaction must be mental, an even more controversial claim follows: namely, that the contents of the conditions of satisfaction must be propositional (see Dreyfus 2000, p. 328). However, this is again a relapse into deliberative, purposeful rationality. Hence Dreyfus refines[52] his critique by arguing that perhaps the point is not that Searle is engaging thoroughly in bad phenomenology but rather that it is only a phenomenology of "effortful, deliberate, thoughtful action, like lecturing on or writing about philosophy, and so leaves out the sort of skillful coping one experiences in the flow of sports or in simply finding one's way about in the world" (2000, p. 329). After all, doing something deliberately, with a 'purpose in mind', is also part of human experience. But Dreyfus has it that "although we often engage in what I call deliberate activities, such thoughtful activity is not the only, nor the most basic, way we relate to the world" (*idem*).

---

[51] According to Dreyfus, Searle's suggestion that human agents at first experience meaningless noises is not only bad phenomenology but also bad science: "Developmental psychologists have found evidence that the human fetus already responds differently to the mother's speech from the way it responds to other sounds. This research suggests that there is no sense in asking from the child's point of view how she learns to take as meaningful the acoustic blasts coming out of people's mouths. It seems that meaningfulness does not have to be learned. Rather, the talking that comes out of people's mouths is always already experienced by the child as meaningful, although, of course, the child has to learn the meaning" (1999, p. 13).

[52] The reader should be reminded that this 'give and take' between Dreyfus and Searle, which often repays reconsideration from both Dreyfus and Searle and thus the refinement of mutual criticisms, is due to its being a debate spanning several years and various journal papers and book chapters.

This can be illustrated following a dictum by Merleau-Ponty, to whom Dreyfus often turns for insights into phenomenological embodied experience: "The polarization of life towards a goal is entirely unrepresented. Objective thought bypasses true intentionality, which is *at* its object rather than positing it" (2005, p. 446). It turns out then that "the phenomenological conditions are more basic than the logical ones" (Dreyfus 2001, p. 186). And yet, Searle still holds that whatever characterization is to be given to intentional content, when it comes to intentionality, it need not be sentence-like, but surely though propositional, just like my dog can be said to have intentional states with conditions of satisfaction, which therefore have propositional content. However, "my dog does not think in sentences" (2001b, p. 278). In this regard, Searle's point is that "the logical structure is pervasive whether the activity is skillful coping or deliberate action" (*idem*). As a matter of fact, any habitual activity can be decomposed in its logical constituents. So Searle thinks they must be part of the experience itself, because the agent, asked about what she just did, resorts many times to set-by-step explanations or to the rule-following she was given when being a novice performer. But whether this is originary is, of course, a controversial matter.

For Dreyfus, however, absorbed coping is more primordial than *a posteriori* reflection on experience, as it constitutes "the background condition of the possibility of all forms of comportment" (1999, p. 11). A panoply of culturally accepted practices lack any linguistically describable status. For instance, distance standing—an example Dreyfus is rather fond of—cannot be reducible to any explicit rule-like structure nor with recourse to any measurable *physical distance* that one could take into consideration. For example, when I see that my child is inappropriately standing too close to someone while standing with me in a queue at the bank, I need only correct his position which I, at the moment, consider too close. In cases such as these, an agent who knows how to cope with such uncomfortable situations as being too close to someone "need only be skillfully moving to lower a tension" (1991, p. 17). There is no written rule about it, we just know how to act in such cases based on cultural experience. Social norms of this sort are too specific, given that they are context-bound and, to that matter, indexical, and therefore cannot be easily described in propositional or representational terms. Such norms are, indeed, indexical because they are produced by their concrete situational conditions of existence (Dreyfus 1999, p. 20). According to Dreyfus, this brings one back to

phenomenology, precisely because human agents are always already in the world. There is then no need to resort to the strange strategy of making intelligible a meaningless world by bestowing meaning upon it from the outside. This is the so-called primacy of phenomenology over logical analysis: not that logical analysis is forever futile and should be barred from our consideration of reality, but that a meaningful world is first and foremost disclosed to us. It is "a kind of third being that is neither natural nor constituted, but is produced by the embodied intentionality that is always already present in the world of involved, active, social beings" (Dreyfus 1999, p. 21). But still, these reasons turn out to be unfathomable for Searle:

> [Dreyfus] says, for example, that when people move to a comfortable distance from other people in an elevator, they do so *unintentionally*; they have no intentions. I do not think that can be a correct description. This is a typical case of intentional action. It is not premeditated; there is no prior intention. And it may be done without even the agent's awareness that he is doing it, but all the same, it is not the peristaltic contraction of the gut. It is clearly intentional. (2005, p. 334)

This unnerving discomfort by Searle is due to his theoretical leanings, specially his conception of intentionality as purposeful and willful action. As we are about to see, however, Dreyfus is after the background of the foreground with his emphasis on nontheoretical, absorbed coping.

Intelligence Without Representation

It may be worth remarking that Dreyfus's nontheoretical account of human coping is structurally isomorphic with a nonrepresentational account of learning standing over against a representational conception of knowledge. Dreyfus is thus congenial to the approach according to which learning by experience need not resemble a trial-and-error process that makes inferences from past interactions represented in the mind—let alone from the association of memorized ideas needing to be retrieved from a list that storages past events. The relevant question to be raised on this regard is how human agents organize experience based on single cases, that is, how generalization from particular interactions with the world occurs. Moreover, by virtue of what is it that human agents find similarity or dissimilarity between situations so that they can acquire the skills needed to get around in the world? On Dreyfus's account, contrary to what a GOFAI-oriented approach to learning would have us believe, similarity cannot be a simple case of memory retrieval from previous inputs. As the case of the connectionist network mentioned earlier shows—which rather clumsily generalized the difference

between a forest with and without shadows (see Dreyfus 1992, p. xxxvi)—"the similarity problem at its most basic is not a question of comparing representations" (Dreyfus 2002a, p. 375). Dreyfus has it that this can be explained as follows:

> One doesn't see the current input as similar to a remembered previous input. One sees it as an input that is impoverished *in a particular way*, and in one's perception of the input one already knows what would count as a better version of it. In this way the perceiver solves the problem of the circularity of associationism by seeing the input as a deviation from the prototypical input for a given object, and so doesn't have to associate the input to the output on the basis of some arbitrarily chosen similarity. The difference is between seeing a garbled input and then guessing what previously learned input it most resembles, on the one hand, and seeing a given input directly as a deviation from a prototypical input, on the other. (*idem*)

No comparison between representations is therefore being carried out at the basic level of absorbed coping. How, and by virtue of which aspects, human agents are able to generalize and find similarities, argues Dreyfus, is due to brain architecture and their actual bodily structure. There are perceptual constants, like the ones proposed by Gestalt psychologists (seeing *Gestalten* or visual scene configurations like proximities, similarities, and continuities, for instance) constrained by innate brain structures, which are given from the start.[53] But this would not be enough if human agents were not embodied beings: "for example, things nearby that afford reaching will be noticed early and often. Their various ways of being reachable and the kind of grip they provide will be an obvious source of shared similarities" (Dreyfus 2002a, p. 376). So Dreyfus's often noticed keenness for neural networks should not be exaggerated.[54] As a matter of fact, neural networks are at a serious disadvantage when it comes to learning to cope with the human world, because "nothing is more alien to our life-form than a network with no up/down, front/back orientation, no interior/exterior distinction, no preferred way of moving, such as moving forward more easily than backwards, and no emotional response to its failures and successes" (Dreyfus 2002a, pp. 376-377). The odds are hence extreme against a disembodied neural network being able to cope with the world the way human agents do (primarily, of course, because a neural network is not in a situation, lacking both a world and existential

[53] As is widely known, Gestalt psychologists "were concerned with how the overall configuration of a scene, rather than particular elements in it, informed the interpretation of the scene" (Bennett & Hacker 2013, p. 10).

[54] Clark notes, for example, that Dreyfus's tilt towards neural networks gives the impression of falling short of a clear demonstration that internal representations of some substantial kind are not involved in expert performance, "since many fans of such networks (e.g., Paul Churchland, myself and others) routinely depict them as supporting a new and powerful kind of internal representation viz. high dimensional weight-space encodings of the prototype structures that characterize a domain: encodings that yield (in the context of a given input) a distributed representation appropriate to the control of action or response" (2002, p. 386). These and other objections of like nature will be dealt with in due course.

understanding). Disembodied creatures, as Dreyfus asserts, still look to us hopelessly stupid (2002a, p. 377). The point, for Dreyfus, is that disembodied and nonembedded creatures, which therefore lack existential understanding, cannot be really counted as intelligent.

What are then the requirements accompanying real intelligence? Appropriating a concept from Merleau-Ponty, Dreyfus thinks that what is needed is a 'maximal grip' which makes an action possible without representing a goal: "according to Merleau-Ponty, in absorbed, skillful coping, I don't need a mental representation of my goal. Rather, acting is experienced as a steady flow of skillful activity in response to one's own sense of the situation" (2002a, p. 378).[55] This sense of the situation is exemplified by many daily actions and motor intentions like seeking a better standing position in an art gallery or grabbing an object. Human agents tend to get the best grip on things and this is constituting of how the grasping of something as something and bodily understanding occurs. The maximal grip is indeed not something that one could represent to oneself: "one only senses when one is getting closer or further away from the optimum" (Dreyfus 2002a, p. 379); nor is it something that one could normally express. As argued previously *contra* Searle, skillful coping does not require mental representation, for actions can be purposive without the agent entertaining a purpose, and *a posteriori* reflections upon how one does it, tend to deform the very experience of doing it, mainly because this amounts to an objectification of meaningful action.

All this, however, still looks like magic for the representationalist theorist of mind, and Dreyfus thinks it appropriate to found phenomenological notions like the maximal grip and the intentional arc in the brain activity underlying nonrepresentational coping by resorting to Walter Freeman's attractor theory; the central idea being that this neurological theory helps to thwart a representational and propositional understanding of behavior. On this account, the sense of the situation, tending towards the maximal grip by relieving the tension of a deviation or, quite the contrary, sensing that one's situation actually deviates from some optimal body-environment relationship, can all be explained by this theory of brain activity.

---

[55] Merleau-Pontyan quotes provided by Dreyfus on the maximal grip are taken from the *Phenomenology of Perception*. See, for example, Merleau-Ponty 2005, p. 250, p. 153 and p. 302.

It should be noted that Dreyfus's objections directed towards Searle's contention that the logical structure is pervasive, inescapable, and encompassing with regard to human activity, are here complemented with what he calls the 'avant-garde' neuroscience of Walter Freeman (see Dreyfus 1999, p.11; Freeman 1991 & 1999). Given that, from the physiological point of view, perceptions are supposed to be created by anticipatory neural activity emerging in the brain, the most important problem for neurobiologists is to render an account of the synaptic mechanisms by which neurons interacting in massive numbers give rise to patterns of anticipatory neural activity controlling behavior. Thus, for Freeman, the biology of brains should be approached in terms of how they construct intentional behavior, and this opposing both the behaviorist stimulus-response and the cognitivist rule-driven logical paradigms (2008, p. 231). On Freeman's view, brains are chaotic dynamic systems, where chaos is not deterministic, but stochastic, that is, "it arises from and feeds on the randomized activity of myriads of neurons, and it provides the basis for self-organization" (2000, p. 13). Freeman (1991) explains the difference between chaos and randomness by distinguishing the behavior of commuters dashing through a train station at rush hour and the behavior of a large, terrified crowd. Whereas the former, although chaotic, is ordered because everyone is hurrying to catch a specific train, the latter is mass hysteria, which is random. This might be, according to Freeman (1991), the chief property that makes the brain substantially different from an AI machine, and it might be as well what confers on the brain the possibility of continually producing novel activity patterns. The brain, says Dreyfus, "can thus be understood as a dynamical system with energy peaks and valleys" (2002, p. 382), so that neural connections are formed when perception leads to a burst of global neuronal activity on an specific energy landscape. On this account, such patterns can be devised as the result of neuronal bursts that resemble contour diagrams indicating the elevations of mountains and valleys (Freeman 1991). Dreyfus goes on to explain Freeman's theory, based on his experiments with rabbits' olfatory system, as follows:

> Past experience has set up the neuron connections so that the current perceptual input, which is similar to some past input but never exactly like it, puts the brain area that controls movement into a specific energy landscape. Once that brain area is in that landscape, movements are caused that tend to move the brain state closer to the bottom of the nearest basin of attraction. The rabbit, or in my example the tennis player, presumably senses this tendency of the system to seek a minimum energy state as a tension drawing it towards a optimal gestalt or maximal grip. [...] The system thus redirects the player to make those movements that result in the brain approaching the lowest accesible point in its current energy landscape, without the player needing to represent where that lowest point is or how to get there, any more than a river needs to

> represent as its goal the lowest point in the landscape in order to find the optimal path down a hill. (2002, pp. 382-383)

This is very important for Dreyfus, because consequently it is thanks to Freeman's work that Merleau-Ponty's claim that the representationalist philosophy of mind is mistaken, "can be defended not only on the phenomenological level but on the neurological level as well" (2002, p. 383).

However, this is not clear to everyone. Clark has replied that it remains an ambiguity on Dreyfus's account, "where mind, cognition and intelligence stop and the non-mental world begins" (2002, p. 385). So the essential question to be asked, for Clark, is where one is to locate the cognitive/noncognitive divide itself and upon what grounds is a line to be drawn (2002, p. 387). What is more, for Clark, Dreyfus's insistence in the nonrepresentational character of mental processes "falls suspiciously short of a clear demonstration that internal representations of some substantial kind are not involved" (2002, p. 386). A similar case against Dreyfus has been brought forth by Rey (2002), who goes on to argue that "it's enough for representationalism merely that some mechanisms of mind are representational" (p. 403), whereas, on the contrary, Dreyfus seems to be arguing that none are. Since, for Rey, representation seems to be the very essence of mind (*idem*), the sole suggestion that brains respond to solicitations without representing them as such, is at the very least a hard idea to grapple with. After all, it might seem pretty obvious or uncontroversial that human minds represent features of their environment (Jackson 2002, p. 409 ff.), or that at least some mental states have representational content (Tye 2011, p. 253). A more fundamental difficulty is endemic to Dreyfus, who might be extrapolating consequences from his well-known dissatisfaction with AI: "the problem with traditional AI was not its representationalism, but its vision of representations as logical symbol structures, and its conviction that intelligence is best understood as problem solving via manipulation of such symbol structures" (Grush & Mandik 2002, p. 389). On this suggestion, Dreyfus would be advancing a dichotomy between skill and representation, which is wrong-headed, for "much skillful performance requires counterfactual reasoning, where one is not letting the world be its own representation" (Grush & Mandik 2002, p. 393). This is the case when, for example, human agents entertain nonactual scenarios or imagine different possibilities: what are they doing in such cases if not representing what is actually not there in the world? In the final analysis, it might be the case that even the intentional arc and the maximum grip themselves, which Dreyfus presents as conceptual options undermining

representationalism, involve representations (Antony 2002, p. 396). Even if the expert has successfully automated her motor routines and has forgotten about how she does cope with certain features of the world, which in previous stages of her learning appeared as salient, this "hardly shows that representations were not a crucial part of the process" (Antony 2002, p. 399).

However, Dreyfus is not entirely repulsed by the idea that there are some processes that can be deemed representational. The question is not "whether *much* skillful performance requires counterfactual reasoning but whether *all* of it does" (Dreyfus 2002b, p. 414). Moreover, on Dreyfus opinion, the suggestion that there is something inside human agents that solicits their actions, that is, inner mental states, "could well be a *body-set* based on how my brain has been modified by my past experience. It is not necessarily a *mental representation*" (2002b, p. 415). It is for want of a better theory that cognitivists so often resort to the argument that the automatization of motor routines allows for the consciousness of representations to go away, which does not amount to the entire disappearance of them. On this view, this means that representations are there, whether one is conscious of them or not. This line of reasoning has led Dreyfus to mock the traditional view that representations are exerting their mysterious powers in human agents, even if the latter are not conscious of their workings. In fact, "by parody of reasoning, one could argue that, since beginning bicycle riders can only stay upright by using training wheels, we should conclude they must then be using invisible ones, and the burden of proof is on anyone who thinks otherwise" (2002b, p. 416). At this point, it is important to underscore the gist of Dreyfus's argument: "on-going absorbed coping produces the intelligibility and familiarity on the basis of which all action is possible." (2002b, p. 418). Therefore, "in the last phenomenological analysis, absorbed coping, as the background condition of possibility of all forms of comportment, makes planned behavior possible and not the reverse" (*idem*). So it is not the other way around, for even if there are representations and propositional knowledge, this kind of knowledge is only derivative and not original. There is, indeed, propositional knowledge, and Dreyfus is not in the business of denying it, but it does not stand alone and, most importantly, it cannot render an explanation of that basis of knowledge, which Dreyfus calls absorbed coping. But is this really the case? Is there something nonpropositional at the very basis of propositional knowledge? In lieu of just denying that all coping is representational, Dreyfus quarrels with the idea that deliberative and propositional reason are to be

given primacy in explanation. But this is so far merely skeletal and hanging further flesh on it requires a discussion of the place of the conceptual in knowledge, which is provided in the very interesting debate between Dreyfus and McDowell.

| 5 |

THE BATTLE OF MYTHS

The Dreyfus-McDowell Debate

Dreyfus's crucial idea for denying so vehemently the preeminence of explicit, theoretical knowledge is beholden to his belief that the primacy of phenomenology over deliberative reasoning amounts to its apriority and pervasiveness. To put it bluntly, absorbed coping precedes—and is therefore more originary than—deliberative reasoning. In fact, no suitable description of the conceptual upper floors of the edifice of human knowledge can succeed in absence of a fitting understanding of the embodied, absorbed coping going on on the ground floor (Dreyfus 2005, p. 47). The same problem applies, of course, to GOFAI and expert systems with their failed attempts to answer the enigma of human knowledge by tackling the issues as from above, that is, by programming theoretical, explicit knowledge in the assumption that such an endeavor will be on the right track for a thorough understanding of expertise and intelligence. The upshot of the cognitivist tradition is marked by the importance it bestows on conceptual activity, even to the point—as Searle's ideas have shown—of conceiving of matters of perception and skillful coping as denizens of an assigned preeminence of logical conditions of satisfaction.

Dreyfus has noted that McDowell is prone as well to this primacy of the conceptual, even when McDowell is known for his belief that "an experiencing and acting subject is a living thing, with active and passive bodily powers that are generally her own; she is herself embodied, substantially present in the world that she experiences and acts on" (2000, p. 111). On McDowell's explanation of human experience, understanding the content of perceptual experience as nonconceptual is an unqualified claim, which is based on the idea that, when referring to experience by means of a judgement, "one moves from non-conceptual knowledge to conceptual content" (2000, p. 47). Certainly, this way of conceiving things reminds one of Kant's dictum that intuitions without concepts are blind, and this is precisely McDowell's Kantian objection to any account of perception pretending to get rid of

conceptual content (2000, p. 53-54). Moreover, "to say that an experience is not blind is to say that it is intelligible to its subject as purporting to be awareness of a feature of objective reality: as a seemingly glimpse of the world" (2000, p. 54). This is nonetheless a concession on McDowell's account to what Dreyfus calls the "Myth of the Mental" (2005, p. 52): the idea that all intelligibility, even perception and skillful coping, must be conceptual, even implicitly (2005, p. 51). On this regard, Dreyfus plays the Merleau-Pontyan objection to intellectualism on McDowell's all-pervasive conceptual powers: "for the intellectualist, judgement is everywhere pure sensation is not, which is to say everywhere [see Merleau-Ponty 2005, p. 34]. For McDowell, *mind* is everywhere the pure given is not, that is to say, all the way out" (*idem*). This talk of 'the given' refers, of course, to Sellars's well-known critique of the myth of the given (see 1997), which plays a major role in McDowell's theoretical framework. Dreyfus foists the Sellarsian idea that "perception is conceptual all the way out" (2005, p. 47) on McDowell, and names it 'the myth of the mental,' meaning that mindedness is all pervasive.

So for McDowell, concepts are somehow supposed to be playing a role in the whole of perception, even if one is not conscious of how this happens. Dreyfus's point is that this was also an assumption being entertained both by GOFAI practitioners and knowledge engineers alike: in some way, rules and concepts become unconscious when expert knowledge arrives. However, repeating the training-wheels parody as objection, Dreyfus inveighs against this assumption that "our experience suggests that rules are like training wheels. We may need such aids when learning to ride a bicycle, but we must eventually set them aside if we are to become skilled cyclists" (2005, p. 52). Moreover, adds Dreyfus,

> to assume that the rules we once consciously followed become unconscious is like assuming that, when we finally learn to ride a bike, the training wheels that were required for us to be able to ride in the first place must have become invisible. The actual phenomenon suggests that to become experts we must switch from detached rule-following to a more involved and situation-specific way of coping. (*idem*)

Against this cognitivist assumption that ascribes theoretical primacy to rule-following in coping (by means of which instructions can then be internalized as if stored in the mind), in habitual human experience "the learner develops a way of coping in which reasons play no role" (Dreyfus 2005, p. 53). When an expert is required to give an account of how she does it, she is then forced to render a reasoned explanation of how what she does leads to her action being accomplished, but this account

involves necessarily "a *rationalization* that shows at best that the expert can retrieve from memory the general principles and tactical rules she once followed as a competent performer" (Dreyfus 2005, p. 54). However, this is an exhibition of competence, but not of performance. In order to support this point, Dreyfus touches upon Heidegger's idea that in performance what is first and foremost given is "the 'for writing,' the 'for going in and out,' the 'for sitting.' That is, writing, going in and out, sitting, and the like are what we are *a priori* involved with. What we know when we know our way around" (GA 21, p. 144). As a matter of fact, there is something about practical dealings with the world which implies a nonthematic approach. But this nonthematic, nontheoretical approach, does not imply blindness, which seems to be McDowell's Kantian preoccupation that mere intuitions be blind. Dreyfus is thus puzzled about how something given, which is nonconceptual, can be converted into a given with conceptual content (2005, p. 59). Although we share some common qualities with other animals, what makes us special is the fact that "we can *transform* our unthinking non-conceptual engagement, and thereby encounter new, thinkable, structures" (2005, p. 60). That is to say, "our ground-level coping opens up the world by opening us to a meaningful Given—a Given that is *non-conceptual* but not *bare*" (2005, p. 55). Only because the ground floor of absorbed, nonconceptual, and thus nonthematic coping constitutes our basic being-in-the-world, it is possible to form beliefs, make judgements, justify inferences and the like, which means that this ground floor is way more basic than the upper floor of theory. Hence Dreyfus's insistence that, whatever mindedness is, it surely grows out of being-in-the-world and not conversely (2005, p. 61). In this sense, the coper does not even have to pay attention to what she is doing, let alone have a theoretical sense of the situation. Instead of conceiving of the background as constituted by mental representations, theoretical observations, and monitoring attitudes, Dreyfus urges us to think of it as a space of motivations calling the coper's attention to act upon a situation by being solicited in order to get a grip on what is currently going on (2005, pp. 56-57). Therefore, absorbed coping can be best described by drawing from Gibson's concept of 'affordance,' for it allows Dreyfus to think of "this conceptually pure yet meaningful given" (2005, p. 55). According to Gibson, "an affordance cuts across the dichotomy of subjective-objective and helps us to understand its inadequacy. It is equally a fact of the environment and a fact of the behavior. It is both physical and psychical, yet neither. An affordance points to both ways, to the environment and to the

observer" (1986, p. 129). This is Gibson's crucial idea: rather than perceiving determinate objects and then represent them as affording certain possibilities for action—as traditional cognitive science would hold—what we actually perceive is the affordances themselves: the possibilities for action. On the traditional (cognitivist) view, perception is categorizing and classifying, but as Nöe has argued: "Gibson's own theory of affordances was advanced as an alternative to this perception-as-classification idea. This is what he had in mind when he said that we see affordances directly" (2013, p. 183).

Let us now turn to McDowell's view on the matter, in agreement with which experience in rational animals, like human beings, is pervasively informed by the capacities that belong to rationality. Accordingly, "something similar holds for our intentional action" (McDowell 2007a, p. 338). Knowledge is thus conceptual all the way right down to intentional action and perception. This is why even unreflective bodily coping is informed by rationality (*idem*), which means that embodied coping must be considered as permeated with mindedness (2007a, p. 339). This is why Dreyfus claims that McDowell finds rationality and mindedness everywhere. On McDowell's account, however, this should not mean that rationality is essentially detached from particular situations, as if mindedness could only mean abstraction from situated knowledge. By this is meant that there need not be any essential connection between rationality and situation-independence. On McDowell's understanding of practical rationality, "affordances are no longer merely input to a human animal's natural motivational tendencies; now they are data for her rationality" (2007a, p. 344). McDowell does not deny the existence of affordances, solicitations to act upon situations, or skillful coping. But they are nonetheless intelligible, and hence rational. Moreover,

> our relation to the world, including our perceptual relation to it, is pervasively shaped by our conceptual mindedness. An implication of this for perceptual content can be put like this: if a perceptual experience is world-disclosing, as opposed to belonging to the kind of coping with a mere environment that figures in the lives of creatures lacking orientation towards the world, any aspect of its content is present in a form in which it is suitable to constitute the content of a conceptual capacity. (2007a, p. 346)

As far as McDowell is concerned, when perceiving, the coper is not involved in mere irrational dealings, but perceiving itself is world-disclosing and therefore actively intelligible. Given the fact that the perceiver can reflect upon what she is doing and considering as well that the very act of perception is intelligible when experienced, why is one to think that skillful coping is somehow not pervaded with rationality? What is irrational about it? What is mythical about it? For McDowell "the real myth in the

neighborhood is the thought that makes it look as if affirming the pervasiveness of conceptual rationality will not cohere with giving proper weight to the bodily character of out lives" (2007a, p. 349). This is the "Myth of the Disembodied Intellect" to which Dreyfus, on McDowell's account, falls prey (*idem*).

Dreyfus's Phenomenological Pitfalls

In the subsequent responses belonging to this heated debate (see Dreyfus 2007b; McDowell 2007b; Dreyfus 2007c; Dreyfus 2013; McDowell 2013), the differences laid bare deepen even more, with Dreyfus defending the claim that conceptual articulated mindedness is actually the enemy of embodied coping (2007b, p. 353). Although lamenting having foisted beforehand the idea of a disembodied intellect on McDowell, Dreyfus sticks to the opinion that mindedness, "far from being a pervasive and essential feature of human being, is the result of a specific transformation of our pervasive *mindless* coping" (2007b, p. 353). Perception is not primarily conceptual because the world, as the background of experience, is not to be understood as a whole of interconnected facts and propositional attitudes one can resort to, but instead as "the totality of interconnected solicitations that attract or repulse" (Dreyfus 2007b, p. 357). When coping in the world, the coper finds herself caught in a web of attractions, repulsions, and solicitations to act upon what needs to be pursued and carried out, not within the context of propositional attitudes and capacities to step back and monitor activity. Far from it, an explicit propositional take on things is but an exception in everyday coping. Our relation to the world, adds Dreyfus, is "more primordial than our mind's being open to apperceiving categorically unified facts" (Dreyfus 2007b, p. 359). As a matter of fact, "this objective world and its conceptual order presupposes a preobjective/presubjective world" (Dreyfus 2007b, p. 360). The ground floor, whose capacities are fundamentally heterogenous with reason (see Dreyfus 2005, p. 47), is constituted by a pre-linguistic horizon which is always presupposed when a minded attitude arrives.

However, if the discussion is framed along similar lines—Dreyfus claiming the powers of a nonrational relation to the world—it could lead one to confusion, as if being-in-the-world were some

kind of irrational stance.[56] So the question concerning rationality must be raised. In this regard, the question can be formulated as follows: are we essentially rational animals? (see Dreyfus 2007d). The discussion between Dreyfus and McDowell can be summarized in that the former answers no, whereas McDowell asserts that rationality is what constitutes human being in its very core. Schear (2013) has elaborated on this debate by pointing out that, for McDowell, "experience, *so far as it matters for rational knowledge of objective reality*, cannot be nonconceptual" (p. 287), that is to say, rationality turns out to be the proper background of the foreground of human activity. And being rational means to be endowed with a set of conceptual capacities. Rational (conceptual) capacities in this manner are 'in play,' or 'permeate' or are 'operative' in our intelligible human activity (Schear 2013, p. 290). Dreyfus seems to construe this permeation of rational capacities brought forward by McDowell as the constant exercise of an abstract monitoring perspective on activity, which causes a subject/object fission and thus a change of attitude in coping. Besides, Dreyfus might have in mind the mental representation view of concepts (that they are psychological entities), which is the default position in cognitive science and the philosophy of mind (see Pinker 1994; Carruthers 1996; Margolis & Laurence 1999).

But it could be argued against Dreyfus that capacities are not exhausted by their being exerted on particular occasions. On the contrary, it could be that they are pervasive inasmuch as they are at the same time general. I might have the capacity to jump, or do pirouettes, or even be kind, which proves that capacities can be well conceived of as pervasive because they remain dormant, so to say, and are only activated and put into practice whenever is needed to exercise them as such. From this follows that this pervasiveness "does not entail the constant de facto exercise of the capacity" (Schear 2013, p. 291). By the same token, that human agents are characterized by rationality—the latter being a pervasive feature of their lives—does not entail necessarily that an abstract capacity is being exerted while practical activities are in play. This is indeed a weak reading of the 'venerable thesis' (the thesis that which ascribes rationality to human agents as their most essential feature, see Schear 2013, p. 285), since being endowed with rationality is not the same as claiming that humans are always rational, nor that they are always monitoring their coping activities, nor even that they are constantly aware of what they

[56] Although Dreyfus seems to ignore it, his idea that there is an essential human dimension which is not entirely rational might ensue from a rather traditional identification between life and the irrational. This idea has cavorted the pages of transcendental philosophy since Kant (see Baeumler 1967) and it might be due, as Molina suggests, to the ambiguity of the concept of life that one can find already in Kant's philosophy (see Molina 2010).

are doing. It should be apparent that this detached idea of rationality is not the one which McDowell has in mind when conceiving human agents as rational beings. In fact, rationality can only be pictured as disrupting the flow of coping in the world if it is simultaneously understood as exercised in a stepping-back, monitoring attitude towards what is being done. However, "the question is whether rationality *qua* capacity is pervasive, where its pervasiveness does not consist in the process of an ongoing constant exercise (which would indeed be inimical to flow)" (Schear 2013, p. 292). So if conceived as a general human capacity, rationality is not to be obfuscated by reducing it to a mere monitoring observation of one's own activity. This is admittedly McDowell's central claim against Dreyfus, or against what he has called 'the myth of the mind as detached', to which Dreyfus, on McDowell's view, succumbs:

> The idea is not that our experiential knowledge is always the result of determining what reason requires us to think about some question. Normally when experience provides us with knowledge that such and such is the case, we simply find ourselves in possession of the knowledge; we do not get into that position by wondering whether such and such is the case and judging that it is. When I say that the knowledge experience yields to rational subjects is of a kind that is special to rational subjects, I mean that in such knowledge, capacities of the sort that *can* figure in that kind of intellectual activity are in play, not that a subject who has such knowledge on the basis of experience is in that position as a result of actually engaging in that kind of intellectual activity. (2013, p. 42)

Be that as it may, Schear (2013, pp. 293 ff.) brings to bear another Dreyfusian counterargument against McDowell, which is certainly more interesting and must be discussed in depth: his denial that absorbed coping has conceptual form. This is the so-called argument of merging: the idea that in the flow of skillful activity, the coper cannot be easily distinguished as a subject dealing with objects. On the contrary, as in the often quoted passage from Merleau-Ponty's *The Structure of Behavior*, Dreyfus underscores the merging character of absorbed coping, whereupon, say, a football player becomes rather one with the field: "the field itself is not given… the player becomes one with it" (Merleau-Ponty 1966, p. 168; quoted by Schear 2013, p. 296). Dreyfus's vocabulary to refer to the experience of absorbed coping is decisively nonobjective, because a vortex of forces in which attractions and repulsions constitute the field of activity is not objective in the sense of being there present for theoretical inspection. Conversely, theoretical inspection is possible only on the basis of this vortex of forces that give shape to our bodily and skilled familiarity with the world. As a matter of fact, Dreyfus distinguishes a background coping from a foreground coping: "the familiar background coping can

support a foreground coping in which the 'I do' is operative" (2013, p. 28). For Dreyfus, not even the I, not even subjectivity in the form of an Ego, is operative in the background: such an Ego, it seems, is also too abstract for Dreyfus, whose preference goes to conceiving human beings in primordial coping as substantially involved and absorbed, out of themselves rather than encapsulated in a subjective consciousness.[57] It is indeed a central plank of Dreyfus's description of skillful coping that the background "goes all the way up to engulf the foreground" (*idem*), meaning that absorbed copers are as if under a spell. Merleau-Ponty's words come to mind: "the orator does not think before speaking, not even while speaking; his speech is his thought. The end of the speech or text will be the lifting of a spell" (Merleau-Ponty 2005, p. 209. Quoted by Dreyfus 2013, p. 28). In this way, the merging character of absorbed coping precludes the presentation of determinate objects, since they can only be presented as such—etymologically: can be put before or against the eyes—when a change of attitude not compatible with absorbed activity arises: "only when things are not developing normally and no alternative perspective directly draws the coper to replace the current one, does the coper have to represent a goal and deliberate as to how to reach it" (Dreyfus 2013, p. 30). Hence, for Dreyfus it is futile to do exactly otherwise and foist on objecthood a primacy which is only secondary or derivative, arising from a background whose nature is precisely nonobjective. Dreyfus quotes Heidegger: "precisely in order to experience *what* and *how* beings in each case *are* in themselves as the beings that they *are*, we must—although not conceptually [*nicht begrifflich*]—already understand something like the what-being and that-being of beings" (GA 29/30, p. 519).

However, Dreyfus's idea that absorbed coping would then not be graspable by rationality is controversial, for it is one thing to say that coping is nonconceptual (inasmuch as nontheoretical) and another, to claim that it is irrational! Although Dreyfus takes pains to point out that his approach is bequeathed by the phenomenological tradition, a series of critical questions can be raised, since some of his ideas on phenomenology—appearances notwithstanding—quarrel explicitly with some fundamental phenomenological insights. In this vein, one could ask: is absorbed coping irrational

[57] Zahavi (2013) has written critically against Dreyfus's dismissal of subjectivity and questioned its phenomenological character, since "the level of absorbed coping involves a dimension of self-experience—at least in so far as that level is supposed to be experiential rather than simply a matter of nonconscious automaticity" (p. 326). On Zahavi's view, this absence of subjectivity in Dreyfus's conception of skillful coping makes him sound somewhat like Dennett, "the moment that Dennett reaches the conclusion that our commonsense self-ascription of mental states is persistently mistaken" (p. 322). Dennett's criticisms to phenomenology will be dealt with in depth in due course.

because nonconceptual? What does it actually mean to say that "while the background coping is largely unthought, it is not unthinkable" (Dreyfus 2013, p. 27)? Can it then be thought of with concepts? If not, how? Is there a way to refer to it without at the same time deforming it by means of interrupting its natural flow (which seems to be what Dreyfus would hold)? It remains to be shown that some implications deriving from Dreyfus's treatment of the background and its purported absolute nonconceptual nature, are rather orthogonal to key phenomenological ideas.

Despite Dreyfus's constant insistence to have found inspiration in Heidegger's philosophy, the contrast between rationalism and irrationalism is not always the most felicitous to be adopted, at least from a Heideggerian standpoint. Indeed, according to Heidegger, "when irrationalism, as the counterplay of rationalism, talks about the things to which rationalism is blind, it does so only with a squint" (SZ § 136, p. 136). Indeed, there is a phenomenological dimension, that of experiencing (*erleben*) as the provenance of understanding, "with which one does not know what to do, and for which the convenient title of the irrational has been invented" (Heidegger GA 56/57, p. 117). Heidegger indeed deactivates the efficaciousness of the traditional distinction between the rational and the irrational precisely because it is only brought forth in a theoretical stance, from whence it—and many other such distinctions—emerges. Dreyfus is no doubt right in that, in most cases when being asked about it, the agent that thematizes about her coping activity tends to transform the practical, skillful field, so that it is modified inasmuch as referred to as if from without (and, to a certain extent, by replacing thus lived experience with an *a posteriori* reflection upon it). The holistic background certainly appears bereft of all its meaning and thus impoverished if it is treated as a mere set of facts about objects and its properties. This means that what beclouds lived experience is the adoption of that which the young Heidegger (see GA 56/67, p. 87) designated as the primacy of the theoretical (*Generalherrschaft des Theoretischen* or *Primat des Theoretischen*). But it must be stressed that it is a mistake to make of the background something atheoretical or irrational in the sense of wholly obscure and—as it were—mystical. Heidegger has dealt with this claim by laying out the nonthematic contours of practical behavior:

> 'Practical behavior' is not 'atheoretical' in the sense of 'sightlessness'. The way it differs from theoretical behavior does not lie simply in the fact that in theoretical behavior one observes, while in practical behavior one *acts*, and that action must employ theoretical cognition if it is not to remain blind; for the fact that observation is a kind of concern is just as primordial as the fact that action has *its own* kind of sight. Theoretical behavior is just looking, without circumspection.

> [...] The ready-to-hand is not grasped theoretically at all, nor is it itself the sort of thing that circumspection takes proximally as a circumspective theme. (SZ § 15, p. 69)

It should be observed that this is not the same as understanding absorbed coping as blind, irrational or ineffable, as does Schear: "the field of attractions and repulsions is… literally unthinkable, at least for the discursive intellect, and in a sense then, ineffable" (2013, p. 298). Schear's subsequent conclusion is also dubious: "if Dreyfus's phenomenology of merging is faithful, then he has identified a form of activity that falls outside the reach of our power of rationality and its characteristic material, namely determinate objects fit to figure in reasons" (*idem*). Rather, what Dreyfus has done is identify a field unknown to traditional cognitive science, since the fundamental mistake in AI research was to try to program computers by installing the foreground of constituted objects without any trace of the background that enables this constitution. GOFAI researchers attempted to program the theoretical gaze without realizing that practical behavior has, as Heidegger asserts, *its own sight* (*idem*). And not only that: it is as primordial that the practical has its own sight as the fact that even observation is a kind of practical concern. The distinction between theory and praxis is, on this Heideggerian view, also crafted from the theoretical stance. Moreover, if the background is—as Dreyfus asserts—nonpropositional and nontheoretical, that is, if it is always presupposed in every propositional and theoretical stance, the whole AI enterprise framed along these lines appears as the consequence of a category mistake—as the *Cambridge Dictionary of Philosophy* defines it: an error "in which things of one kind are presented as if they belonged to another" (Blackburn 2005, p. 58). In this case, 'things' whose character is essentially nonobjective being presented as if they belonged to the objective realm of facts. Indeed, GOFAI went astray when the background was purely understood in terms of objective relations.

However, there are two claims espoused by Dreyfus which are hardly phenomenological and they are to be dealt with immediately. In the first instance, let us be reminded of the fact that speaking of lived experience—unlike Dreyfus's much preferred examples of everyday dealings with equipment and skilled coping in sports—is much more ample than reducing the background to mere practical activity. Dreyfus's exclusive interest in Division I of Heidegger's *Sein und Zeit* (see Dreyfus 1991) speaks in favor of understanding his interpretative attempts as a sort of 'pragmatization of phenomenology,' in which skillful coping plays a major role. As Braver (2013) has shown, Dreyfus also applies the emphasis on skillful coping he drew from Division I to the second part of Heidegger's *magnum opus*,

missing thus the fundamental Heideggerian shift of accentuation: "Division II's authenticity [*Eigentlichkeit*] presents something of an *Aufhebung* of Division I's antithesis by marrying coping's engagement to theory's attentiveness. [Heidegger] consistently worries about a familiar behavior's tendency to lull us into autopilot, a state he calls fallenness and consistently connects to the unthematic absorption in the world that is Dreyfus's highest state" (Braver 2013, 146).[58] Heidegger himself warned specifically against this line of construing his philosophy, which reduces it to mere practical everyday activity:

> I attempted in *Being and Time* to provide a preliminary characterization of the *phenomenon of world* by interpreting the *way in which we at first and for the most part move about in our everyday world*. There I took my departure from what lies to hand in the everyday realm, from those things that we use and pursue, indeed in such a way that we do not really know of the peculiar character proper to such activity, and when we try to describe it we immediately misinterpret it by applying concepts and questions that have their source elsewhere. That which is so close and intelligible to us in our everyday dealings is actually and fundamentally remote and unintelligible to us. In and through this initial characterization of the phenomenon of world the task is to press on and point out the phenomenon of world as a problem. It never occurred to me, however, to try and claim or prove with this interpretation that the essence of man consists in the fact that he knows how to handle knives and forks or use the tram. (GA 29/30, pp. 262-263)

It must be noted that while in the context of criticizing GOFAI Dreyfus's recourse to practical dealings is understandable, his subsequent pragmatization of the background causes a rather extravagant and unnecessary split, "where the body is intelligent precisely where the mind is stupid and vice versa" (Braver 2013, p. 152). This conceptual/nonconceptual split is even to be interpreted as "two separate ways of being open to the world" (Dreyfus 2007d, p. 108). The question to be raised is, of course, how do these two spheres connect with each other? Moreover, is Dreyfus introducing the well-known dualism between the object of knowledge and the subject which can never be explicitly known, since referring to it would objectify it? According to Dreyfus's phraseology, this is the problem of explaining "how the *nonconceptual given is converted into a given with conceptual content*" (2005, p. 59). Or, as was formulated otherwise, "how the ground floor of pure perception and receptive coping supports the conceptual upper storeys of the edifice of knowledge" (2005, p. 19). While, for Dreyfus, this fission between a nonconceptual given and conceptual content retains its appeal, a phenomenological (Heideggerian) admonition could be applied to it: "is there really this division and separation…

[58] In point of fact, Heidegger asserts that "that in which concern has *fallen* at any given time is not thematically perceived, not thought, not known, and it is just this which grounds the possibility of an *original reality*" (GA 20, p. 263). Also, in a previous university lecture, Heidegger speaks of a tendency to forget oneself as *Ruinanz* (GA 61, p. 121).

between the given (giveable) and the description? Are we not succumbing here to a deception of language, and in fact a theoreticized language?" (GA 56/57, p. 111-112). As Zahavi has argued in explaining Husserlian phenomenology: "to detach sense and the sensuous (*Sinn* and *Sinnlichkeit*) from each other, to deny the continuity between the perceptual givennes of an object and its predicative articulation, is to make the relation between conceptual thinking and perception incomprehensible and contingent" (2003, p. 29).

In addition, relative to this split between the nonconceptual given and the given with conceptual content or this edifice with two floors, which phenomenology, laments Dreyfus, has yet to connect (2007b, p. 364), there is another objection to Dreyfus's understanding of the background that is also problematic from the phenomenological point of view: namely, his demand that the coping background be untouched by any conceptual reference to it under the risk of being distorted. Isn't Dreyfus proceeding with concepts as well when investigating the scope of skillful coping? Or to put it blatantly: how does he know that 'skillful coping' as conceptual label does not modify and alter skillful coping itself?

Let us elaborate on this further. One of the first objections phenomenology had to face was precisely that the flow of lived experience does not remain the same when reflection enters the stage. Since it can be said that every reflective glance at subjective phenomena is modifying, that is, given that the theoretical eye which is looking at the experiencing itself and which pretendes to access it as it purportedly takes place cannot claim for itself the access to purity of any kind, a science—like phenomenology—claiming a pure entrance to *Erlebnis* might run the risk of objectifying that very experience. Inasmuch as reflection, construed as a theoretical stance presenting objects, is essentially—thus conceived—theoretical, phenomenology cannot claim for itself a nonmodifying glance at immediate experience. A science of immediate experience is simply a wooden iron.

Incidentally, this classical objection against phenomenological philosophy is not obsolete but rather very actual. In this vein, it bears reminding that this was also Searle's so-called 'bankruptcy of phenomenology' criticism: that it deals with how things seem to me here and now in the *immediate present* (see Searle 2001, p. 282): a sort of precarious introspective view lacking scientific importance. According to this criticism, this is precisely what would render phenomenology as a flawed

philosophical endeavor from the very outset of its investigative journey. Thus, the *peccatum originale* of phenomenology would be its false aspiration of providing an access to pure subjectivity, while objectifying subjectivity in the very process.

However, Husserl reacted to these criticisms regarding 'self-observation' and 'immediate experience' in 1913 in order to distinguish phenomenological philosophy from mere introspection. He thus claimed that 'self-observation', 'immediate experience', or what most recently is called 'inner experience' (Price & Barrell 2012) and 'introspective evidence' (Jack & Roepstorff 2003; Jack & Roepstorff 2004) are concepts which are not to be conflated with phenomenology, although this confusion is often entertained nowadays (as, for example, in Gallagher 2003). In § 79 of *Ideen I*, Husserl discusses the objections to phenomenology advanced by the experimental psychologist H. J. Watt (a disciple of Külpe and member of the Würzburg School). Husserl quotes at length:

> It is scarcely possible even to form opinions concerning the way in which one comes to a knowledge of immediate experience. For it is neither knowledge nor the object of knowledge, but something different. One cannot see how a record concerning the experience of experience, even if it has been taken, could be put on paper… But this is always the final question of the fundamental problem of self-observation… It is now customary to refer to this absolute description as phenomenology. (Hua III, p. 152)

According to this well-known line of criticism, phenomenology proceeds by acts of reflection[59] and, by doing so, ignores that that which is reflected on, inasmuch as modified, cannot count neither as pure access to the things themselves nor to immediate experience. What is discussed here is the question of the phenomenological access to lived experience, or to say it with Heidegger of "how experiencing as such is to be had" (GA 59, p. 92). How does one refer to experiencing without at the same time and by necessity distorting it and modifying it? By describing it, doesn't immediate experience show as something else, as something which it essentially is not, as something thus changed by description? Given that phenomenology is descriptive, it must use language, and grasping in words is generalizing. The criticism, of course, has as basis the idea that all language is itself objectifying (see Heidegger GA 56/57, p. 111).

[59] According to Husserl, "die phänomenologische Methode bewegt sich durchaus in Akten der Reflexion" (Hua III, p. 144). Moreover, "das Studium des Erlebnisstromes vollzieht sich seinerseits in mancherlei eigentümlich gebauten reflektiven Akten, die selbst wieder in den Erlebnisstrom gehören und in entsprechenden Reflexionen höherer Stufe zu Objekten von phänomenologischen Analysen gemacht werden können und auch gemacht werden müssen" (Hua III, p. 147).

Nevertheless, at least from a Husserlian standpoint, the line of criticism as to the problematic character of 'self-observation' is self-refuting. The point is, of course, not to take reflection as alien to lived experience (*Erlebnis*), because it is itself part of the stream of experience and it should be apprehended as "consciousness's own method leading to the knowledge of consciousness in general" (Hua III, p. 147). It is always possible to cast a glance of intuitive perception at lived experience (see Hua III, p. 104), to dip into it by means of what Husserl calls a *reflektive Blickwendung* (Hua III, p. 84). Thus, on Husserl's terms, it is absurd to affirm that lived experience is epistemologically guaranteed only in so far as the experiences deriving thereof are given to us in immanent perception, pure and untouched by reflection, or that we can be sure of them only in the actual flow of the present moment, as it would be nonsensical to doubt "whether in the end experiences which pass into the field of vision are not precisely for this reason transformed into something *toto cælo* different from what they were" (Hua III, p. 151). It is self-refuting, argues Husserl, to doubt the significance of reflection for knowledge, because "as he asserts his doubt, he reflects, and to set this assertion forth as valid presupposes that reflection *has* really and without a doubt… the very cognitive value upon which doubt has been cast, that it does *not* alter the objective relation, that the unreflective experience does *not* forfeit its essence through the transition into reflection" (Hua III, p. 155). In sum,

> a *knowledge* of unreflective experiences including unreflective reflections is presupposed throughout, whilst at the same time the possibility of such knowledge is put into question. That happens in so far as doubt arises as to the possibility of making any statement whatsoever concerning the content of unreflective experience and the work of reflection upon it: how far does reflection alter the original experience, and does it not falsify it, so to speak, by converting it into something totally different from what it was? (Hua III, pp. 155-156)

As is widely known, a similar line of criticism against phenomenological philosophy was espoused by Natorp a century ago in his *Allgemeine Psychologie nach Kristischer Methode* (2012), for whom it is not possible to grasp the content of immediate experience (*unmittelbares Erlebnis*) as it is in itself, in its pure *Strom des Werdens* (see 2013, pp. 102-103). Even when this is attempted, the result is always objectifying, just as—adds Natorp—the anatomist fixates, isolates, and strips his specimen of life when turning it into an object (2013, p. 103). One terminates thus the life of the subjective [*man schlägt die Subjektivität tot*], while at the same time pretending deceitfully to pinpoint it (*idem*). And this, because reflection implies a sort of de-living stance on *Erlebnis*. Reflection involving thus a theoretical attitude, that of

looking where one does not normally do so (in the stream of becoming), turns out to be ineffective to grasp lived experience since it cannot but be objectifying. One thinks—deceitfully—that in phenomenological reflection one has accessed the very immediate subjective experience, but what truly happens with reflection is that the access in only to that which is reflected upon. So, *pace* the phenomenologist, we never reach immediate experience but only our theoretical glance at it.

Now, it is widely known that Heidegger has eulogized Natorp as being perhaps the only one to have raised serious scientific objections to the phenomenological method advanced by Husserl (see GA 56/57, p. 101), while at the same time acknowledging that he can "allow Natorpian objections to come up, because they themselves stem from the theoretical standpoint" (GA 56/57, p. 102). Natorp's position is so decisively theoretical, that he even conceives of ordinary representations and prescientific knowledge as already objectifying, as the next passage of his *Allgemeine Psychologie* clearly shows: "Die gesamte auch nichtwissenschaftliche Vorstellung der Dinge ist in der Tat das Ergebnis einer oft schon weitgehenden Objektivierung" (Natorp 2013, p. 196). At any rate, for Heidegger, Natorp's critique of phenomenology is attractive, but not because it dismisses effectively the subtle descriptive powers of phenomenology (which it certainly does not), nor because Natorp's own theory-laden absolutism and reconstructive method constitute an effective alternative to phenomenological philosophy. But rather because, on Heidegger's view, Natorp has adequately shown the 'original sin' of Husserlian phenomenology: its theoretical character. As a matter of fact, Husserl's idea of description in phenomenology deals with acts which were originally not present to oneself, since they were athematically lived, but which now must become part of a theoretical posture [*theoretische Setzung*] (Hua XIX/1, p. 14). This means that for Heidegger, Husserl—for all his acumen and mastery—remains stubbornly in the theoretical stance.

Therefore, the question to be asked is whether this method of reflection "is capable of investigating the sphere of experience and disclosing it scientifically" (GA 56/57, p. 100). According to Heidegger, "in reflection we are theoretically oriented" (*idem*): "we are no longer living the experiences, but looking at them. The lived experiences now become looked-at experiences" (GA 56/57, p. 99), which in the final analysis implies abstraction and a true disloyalty to phenomenology's principle of

presuppositionlessness and to the original Husserlian purpose of disclosing the intuitive realization of abstraction [*intuitiven Vollzug der Abstraktion*] (Hua XIX/1, p. 10).

To be sure, Natorp is not claiming that subjectivity is forever inaccesible to science, although he claims that in reflection one only has a paralyzed and objectified sense of the *Erlebnis*. On Natorp's terms, not even by means of refining our objective grasp of subjectivity, that is, not even by the most strict phenomenological approach, will we be able to dissect the *Erlebnisse* and present them as what they truly are. So the only serious possibility at our disposal is to accept this tragic absolutism of the objective perspective and its inescapable character. Therefore, what is left for philosophy is a neutralization of the effects of reflection, a reconstruction of those elements of the unity of the *Erlebnis* that were turned apart in the objective process in order to merge them and, in a way, restore the *Erlebnisse* in their original state (Natorp 2013, p. 192). For Natorp, there is no question that the dimension constituting pure experiencing will thus remain ultimately unreachable and will only occupy and ideal case for the reconstructive process. But Natorp sees no other plausible way of proceeding but to begin with our objectifications and from there advance step by step towards the recovery of the subjective dimension.

What is interesting is that Heidegger seems to accept Natorp's story (at least partly). Indeed, Heidegger seems to think Natorp is right in his suggestion that the subjective dimension of experiencing cannot be adequately investigated at the behest of reflective description: "Description is destruction, immobilitation of the stream of life, objectification. Because through the description the infinite constitution complex is interrupted, the correlativity of consciousness is disturbed" (GA 59, p. 194). So, from this point of view, "Natorp's critique of the description of consciousness is therefore justified" (*idem*).

Notwithstanding this partial agreement with Natorp, Heidegger reproaches Natorp that his reconstruction method is at the same time a construction, that is, it is also an objectifying procedure which is entirely embedded in the theoretical attitude. So it is hard to see how this method should give us access to the immediacy of subjective experiencing (GA 56/57, p. 107). Moreover, given that Natorp has also claimed that the pure dimension of the immediate (*das Unmittelbare*) is actually hidden from us inasmuch as not given, it is not clear how Natorp could guarantee that his own reconstructive method

would determine it "as it was before the analysis" (Natorp 2013, p. 192; Heidegger GA 56/57, p. 107). This could also be couched in the following terms: if immediate subjective experience is inaccessible, since the mere inspection of it distorts it, how does actually Natorp know that? If reflection distorts the *Erlebnisse*, doesn't this imply that there is actually a dimension of pre-givenness that is distorted precisely by means of theoretical reflection? Given that Natorp denies that this dimension can be known, it strikes one as surprising that the neo-Kantian would argue that reflection distorts it. For in order to claim that something is distorted, one would necessarily have to know the structure of that which is thus transformed. Again, if reflection distorts it, then there is something which is thus distorted. Natorp's argument seems to be muddled and confusing at the very least.

Heidegger's answer to Natorp's reconstruction method is an "originary science of life" (GA 58, p. 233), which begins with the recognition that life is not an object, but neither a subject, for *Subjektivierung* is also a theoretical postulate that deforms life (GA 58, p. 145). So Heidegger's point of departure is not psychology but rather factic-life experience itself. Of course, reflection is not originary but only because there is a self-acquaintance that belongs to experiencing as such: "es gilt, diesen im Erfahren selbst liegenden *Charakter des Vertrautseins mit 'mir'* zu sehen" (GA 58, p. 157). On Heidegger's view, this is why, on the one hand, *es weltet* (GA 56/57, p. 73). On the other hand, "das Bedeutsame ist das primäre" (GA 56/57, p. 73). What is more, some questions could be raised regarding description: "what does it mean that one thing describes another thing? Is description as such a way of connecting things? […] Can one talk about of one thing if there are only things? In this case, there would not even be anything at all; not even the nothing, because with the total supremacy of the material thing, there can be no 'there is'" (GA 56/57, p. 62).

So phenomenology is safeguarded when it abandons the traditional distinctions that one encounters in the theoretical attitude. Indeed, according to Heidegger, "the irrational is an idle name that was invented in order to explain that with which one does not know what to do" (GA 56/57, p. 117). Dasein finds itself already in a world which is meaningful (GA 20, p. 352), which explains why I always find myself affected by this or that mood. For phenomenology then the point is not to begin with a distorting reflection of the irrational, but to articulate factic-life experience from the motivations that spring form life itself. On the contrary, the Natorpian conception of how I find myself is entirely

theoretical, since the I is supposedly just an abstract reflective construct that stands detached to its objectified world-correlate.

Be that as it may, it remains a matter of discussion if Heidegger's hermeneutic intuition (his alternative to Husserl's phenomenological reflection) provides the appropriate access to lived experience.[60] Dreyfus, of course, does not neglect the possibility of referring to embodied, embedded, experience and he resorts to a terminology (lines of force, attractions and repulsions, etc.) when giving an account of everyday skillful coping. Despite that, his frequent parlance of the modification of skillful coping by concepts, and his preoccupation that it be thus deformed, shows a certain similarity to the above mentioned criticisms. But it is even more important to note that Dreyfus's demand that a detailed account be elaborated as to how the transition from the nonconceptual background to the conceptual foreground occurs, is not phenomenological. This demand is made precisely because Dreyfus—to use Husserlian terminology—sees a discontinuity between lived experience and reflection. But even from a Heideggerian standpoint, this Dreyfusian discontinuity is flawed and nonphenomenological, since practical activity is not blind, it has its own concernful involved sight, which Heidegger calls circumspection. Moreover, in order for Dreyfus to maintain this discontinuity of spheres, he has had to mistakenly understand Heidegger, for example, when he foists on Heidegger the following idea: "Heidegger points out that most of our activities don't involve concepts at all. That is, they don't have a situation specific 'as-structure'" (2007c, p. 371). Conversely, the structure of *etwas als etwas* is, for Heidegger, pervasive: "every act of having something in front of oneself and perceiving it is, in and of itself, a 'having' something as something… However, this as-structure is not necessarily related to predication. In dealing with something, I do not perform any thematic predicative assertions" (GA 21, p. 17). Indeed, "that which is disclosed in understanding—that which is understood—is already accesible in such a way that its 'as which' can be made to stand out explicitly. The 'as' makes up the structure of the explicitness of something that is understood" (SZ § 32, p. 149).[61] This sounds strikingly close to McDowell's idea that world-disclosing "is present in a form in which it is suitable to

[60] For Herrmann it does (see 2000). For Zahavi, the contrast between reflective phenomenology and hermeneutical phenomenology is merely artificial: "[hermeneutical phenomenology] remains a reflective enterprise, as long as we simply operate with a sufficiently broad notion of reflection" (2003, p. 170). See also Zahavi (2006) on Heidegger's 'agenda': "his own reasons for wanting to emphasize his own originality vis-à-vis his old teacher."

[61] "… das *ausdrücklich* Verstandene, hat die Struktur des *Etwas als Etwas*" (SZ § 32, p. 149).

constitute the content of a conceptual capacity" (2007a, p. 346). Indeed, there are certain McDowellian ideas that sometimes seem much more phenomenological than Dreyfus's.

Heidegger's treatment of the 'as,' on the one hand, explains his notable dissatisfaction with reflection: it relinquishes the realization of vitality (*Vollzugslebendigkeit*) to the benefit of a mere presence (*ein reines Dastehen*). As Marion has argued on this regard: "it is not at all self-evident that the things in question are given only in the form of their constituted objectification" (1998, p. 2). Heidegger is dismissive of a Husserlian project which he judges as being guided by the traditional and metaphysical question as to how consciousness can become the object of an absolute science. But this guiding question, says Heidegger, is not a Husserlian invention after all but rather "the idea that concerned *modern* philosophy since Descartes" (GA 20, p. 147).[62] On the other hand, in view of the the above said it is clear that some methodological measures are to be strictly undertaken in phenomenology. Phenomenology's tendency not towards constituted objects but rather towards the conditions of possibility of knowledge in general has undeniably given the impression that, inasmuch as an enterprise for the description of nonobjective phenomena, its importance for science is rather questionable. As is widely known, phenomenology requires, for Husserl, "an antinatural habitus of seeing [*Anschauungs*-] and thinking [*Denkrichtung*]" (Hua XIX/1, p. 14), which does not consider the objects but rather the acts that underlie them. The Husserlian measures adopted imply an abrupt suspension of naïve metaphysical opinions (*epochê*), which is actually the first step in entering philosophy, that is, the entrance gate (*Eingangstor*) to pure subjectivity (see Hua VI, p. 260). In addition, a reduction (*Reduktion*) is called forth that will enable the thematization of the correlation between subjectivity and world (see Hua I, p. 61).

Heidegger has also taken methodological measures for coming to grips with the athematical. Its investigation has three elements, which have to be taken into account: (i) *ein Gefragtes* (that which is asked about), (ii*) ein Befragtes* (that which is interrogated), and (iii) *das Erfragte* (that which is to be found out by the asking). That which is asked about (*ein Gefragtes*) is being, or, in Aristotelian terms, "that which determines entities as entities, that on the basis of which [*woraufin*] entities are already

[62] What is more: "Die Herausarbeitung des reinen Bewußtseins als thematisches Feld der Phänomenologie ist *nicht phänomenologisch im Rückgang auf die Sachen selbst* gewonnen, sondern im Rückgang auf eine traditionelle Idee der Philosophie" (Heidegger GA 20, p. 147).

understood" (SZ §2: p. 6). That which is to be found out by the asking (*das Erfragte*) refers to the fact that the question is not what is being (in order not to turn it into an entity and remain in traditional metaphysics), but rather about the *meaning of being*: *à la* Husserl, it is about the conditions allowing entities to appear as such. Could one proceed, however, by means of pure ontological analysis, that is, by categorizing every existing thing? If that were the case, being would be an entity or, in any case, something. If we could purely refer to being, if we could translate it or even just utter *it*, then we would precisely stand in front of a *thing*. But "the being of beings 'is' not itself an entity" (*idem*). The ontological difference itself implies that the fundamental element of the investigation is *ein Befragtes*: "entities themselves turn out to be *what is interrogated*" (*idem*). This means that in Heidegger's thought the ontological is not a thematization of being, or of entities, but an explicit account of the being of entities. Given the fact that being is not at the same time an entity, what must be surrounded are precisely entities, where being indeed announces itself and its meaning. What must be sought, in surrounding and besieging the entities, is the lighting-up of its being. Envisaging entities with a view to its being means that "interpretation does not consist in seeing another being, but in seeing being otherwise" (Marion, 1998: 63). Now, this might seem rather tricky, since being is not an entity and that which is not an entity has always been called nothing. And the nothing: "what else can it be for science but an outrage and a phantasm?" (Heidegger GA 9, p. 106).

Now, with regard to Dreyfus's philosophical contribution to cognitive science, it must be granted that it is enormous, for it has enabled the discussion of these phenomenological topics in an interdisciplinary field prone to a certain species of psychologism (and scientism), both of which will be critically dealt with in due course. Although what is designated by the term Background remains blurry in Dreyfus's exposition and his split between perception and conceptuality, along with the disappearance of subjectivity in his characterization of coping, might turn out to be unphenomenological, the problems raised in his work have inspired a breed of researchers looking to overcome central tenets of cognitivism, a central plank of which is skepticism towards what phenomenology can achieve in cognitive science. So instead of going directly to the work of the researchers who have been influenced by Dreyfus and phenomenology, the next chapter will deal first with some coetaneous flawed characterizations of phenomenology in cognitive science; an exposition

of which will prove necessary to pave the way towards a critique of the computational theory of mind from the phenomenological standpoint. As a guiding thread, the problem of consciousness in cognitive science (and how phenomenology tackles it) will be dealt with immediately in the next chapter.

| 6 |

PHENOMENOLOGY AND THE PROBLEM OF CONSCIOUSNESS

Unfolding the 'Mystery' of Consciousness

The conception that the presence of consciousness in the physical world is to be regarded as a mystery, abounds rather heavily in the literature.[63] Although the concept of consciousness has a long pedigree and its roots can arguably be traced back to ancient philosophy, it has come to be difficult for scientists to accommodate the term in empirical research. Kilian (1970), a German physician and scientist writing in the 1970s, argued that the main difficulty was in the first place defining consciousness. For Kilian, the definition problem was above all due to 'research gaps' on behalf of the empirical sciences and 'missing links' regarding a coherent anthropological synthesis in a theory of consciousness, which was then lacking (p. 103). Filling such research gaps turns out to be the challenge which a scientific theory of the mind worthy of the name has to meet. Given that every mental process has its physical correlate, a fundamental theory of mind must not only deal with brain processes and the description of physical occurrences from the neurobiological and physical standpoints. Most importantly, it must be able to deal with the correlation itself between the physical processes and how they are transformed into thinking, meaningful, sentient, and conscious human experience (see Richter 2007). Thus, this previous parting company with consciousness in scientific circles was meant to avoid explicitly mere metaphysical speculation and idealist explanatory attempts, when a suitable theory was lacking.[64]

However, it is widely accepted that the situation has changed recently in favor of consciousness research. As Searle concedes, "consciousness has again become respectable, indeed almost central, as a subject of investigation in philosophy, psychology, cognitive science, and even neuroscience" (1997, p. xi). What gave consciousness research a renewed respectability? The blossoming of a decisive interest

[63] See, for instance, Searle's chosen book title: *The Mystery of Consciousness*, 1997. See also McGinn 1991; Penrose 1990 and 1995; Tye 1997 and 2007; Blackmore 2005 and 2010; Chalmers 2000 and 2007; Frith & Rees 2007.

[64] According to Kilian, American and English science-oriented psychologists were rather fond of avoiding the term 'consciousness' at all cost: "Im Register des 'Annual Review of Psychology' taucht der Begriff seit 1953 tatsächlich nicht mehr auf" (1970, p. 102).

in consciousness research resided in the studies of attention, memory and imagery made possible by the cognitive revolution. But not without irony, as Kihlstrom suggests, "many of those who might have made a science of consciousness have instead gravitated, knowingly or not, toward a science of the mind that gives precedence to unconscious processes" (1999, p. 174). Kihlstrom (1987) has christened 'cognitive unconscious' the contemporary research in cognitive psychology revealing the impact of nonconscious mental structures and processes on the individual's conscious experience, thought, and action.[65] As a matter of fact, in cognitive science and the philosophy of mind reigns supremely the idea that consciousness, no matter its definition, is but the pinnacle of a series of unconscious processes.

Flanagan (1991) has pointed out several reasons that account for this way of investigating consciousness by turning to the unconscious, which can also be seen as a way of deflating the problem of consciousness by investigating its mechanisms. The claim is that consciousness is not an essence but the by-product of a series of functional processes. Among the reasons that explain this research strategy, it must be remarked that some methodological assumptions which cognitive psychology inherited from behaviorism—especially the emphasis on publicly observable behavior as the mark of the conscious—have had as a consequence the disapproval of private and introspective experience. Here consciousness, of course, is defined basically as awareness on behalf of the human agent of her own mental processes, which can be expressed by means of introspective reports (see Eysenck 1999).[66] A piecemeal approach has also been dominant according to which a thorough understanding of consciousness will emerge only from the study of individual phenomena, that is, in a bottom-up fashion. In addition, there is the more radical idea—typical of computational functionalism—that in the final analysis consciousness is rather inessential, since an adequate description of human information processing can be given in terms of the relations between stimulus inputs and response outputs without resorting to phenomenal experience. Finally, among this general dismissal of

[65] A series of experiments have decisively shown the role played by unconscious psychological processes in perception, memory, and action. Unconscious processes were mostly observed in patients with brain damage: "Some patients with lesions in visual cortex can make correct 'guesses' about the properties of visual stimuli that they cannot 'see.' Patients with dense amnesia can retain knowledge about stimuli they have no memory of having seen before. Patient DF, with damage to inferior temporal cortex, can use visual information of which she is unaware to guide her movements" (Frith & Rees, p.16). Frith and Rees take this experimental evidence from Weiskrantz & Warrington 1975; Warrington & Weiskrantz 1968; Goodale et al. 1991, respectively.

[66] This definition, which conflates consciousness with introspection, is rather a disputable one and will be dealt with critically later.

consciousness, there is also an epiphenomenal suspicion: even if consciousness is essential, it is certainly nothing but the end product in a chain of cognitive processes, so it can be argued that it has no causal role in human comportment. The latter renders human agents as mere conscious automata, because if consciousness can be ultimately reduced to unconscious processes, the latter are to be given explanatory priority.[67]

Given that human agents are most of the time unaware of those processes influencing their behavior, Nisbett and Wilson (1977) concluded that conscious awareness does not provide access to useful information about one's own cognitive system. This was also Dennett's early idea that human protocols and human operations (protocol statements) were logically isomorphic and, by the same token, irrelevant for giving an account of machine traces and computer operations (Dennett 1968).[68] Consciousness has been in part so uninteresting for the science of mental processes, that the entry on 'consciousness' provided in *The Blackwell Dictionary of Cognitive Psychology* concludes somewhat dismissively that "it is often important for cognitive psychologists to obtain evidence about the contents of consciousness" (Eysenck 1991, p. 85). *Often*, but not *always*.

As can be seen, the new interest in consciousness research has not precluded complications associated with the term to surface, one of which is ironically the attempt to explain consciousness but only by explaining it away. Sometimes the appearance of the term has even tended to jettison the scientific discussion by relapsing into metaphysical conundrums. As it turns out, consciousness is often seen as a stalemate for its resistance to mechanistic and naturalistic explanations, the scientific question being in this case: how does consciousness emerge from matter? Moreover, how to reconcile the third-

---

[67] For cognitive psychologists with Strong AI theoretical leanings, it is fundamental that agents can be so described, that is, as not being aware of their own mental processes. In PSI, Dörner's (1999) theory for explaining how the mind is entirely possible as computational activity, a computational agent is described as follows: "it [the PSI agent] would not know about the computations that lead to PSI's perception of hunger. The respective processes would remain hidden to it. PSI would just have a motive, 'hunger,' and this would press into the foreground, would get to direct actions, or fail to do so" (1999, p. 806). On Bach's view, "Dörner's PSI agents… are tremendously simplified projections of underlying (and sometimes intangible) theoretical ideas. Currently they do not know grammatical language (which is a tenet of the theory), they have a limited perceptual access to their world (which is a virtual one), and they lack most self-reflective abilities. And yet, Dörner claims that they are already autonomous, know real meaning, possess real motives, and undergo real emotions. Furthermore, it might be possible to extend them along the suggested lines into a full-blown constructionist model of human emotion, cognition, behavior, and personality" (2009, p. 62). This, again, exemplifies a piecemeal approach attempting to construct consciousness in a bottom-up fashion and that assumes as fundamental the fact that agents do not know of their own cognitive operations.

[68] Of course, let us not forget the fundamental cognitive assumption that "the brain is a machine, a biological machine, and it can think" (Searle 1997, p. 13).

person perspective offered by the causal, mechanistic worldview delivered by biology, chemistry and physics with the first-person perspective characterizing human experience? How is it, and why, that a series of impersonal processes give rise to qualitative, phenomenal and personal experience? To use Tye's words, "how do objective, physical changes in the brain generate subjective feelings and experiences? What is the mechanism which is responsible for the production of the 'what it's like' aspects of our mental lives?" (2007, p. 27). This is the so-called hard problem of consciousness (Chalmers 1996), which has the character of phenomenal experience at its core. Following Edelman (2008, p. 412), the hard problem can be posed in the form of three related questions: What makes a 'mere' physical process an experience for someone? What makes a 'mere' physical system a subject, or an experiencer? What does having a first-person perspective on an experience consist of?

For Chalmers, rendering a satisfactory account of the aforementioned questions is very difficult. Consciousness would indeed be the biggest mystery still remaining, perhaps it "may be the largest outstanding obstacle in our quest for a scientific understanding of the universe" (1996, p. xi). The so-called 'hard problem of consciousness,' as Chalmers (2007) has recently redefined it, amounts to the problem of experience and to the subjective aspect accompanying it: "it is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises" (2007, p. 226). The problem of consciousness is indeed hard if one compares it with the rather 'easy' problems.[69] Its stubborn presence suggests that new theoretical ventures will have to be devised if the mystery is somehow to be suitably met (for instance, using chaos and nonlinear dynamics in a comprehensive theory, devising nonalgorithmic processing methods, awaiting for a breakthrough in neurophysiology, designing a quantum mechanical explanation of brain processing, etc.). On this same vein, Penrose has argued that in order to come to grips with consciousness, "we ought indeed to come to terms with the mysteries of quantum theory" (1995, p. 235). However, Chalmers argues that quantum mechanics retains its appeal because its attractiveness may stem from a Law of Minimization of Mystery: both consciousness and quantum mechanics are puzzling, so the mystery might have a common source (2007, p. 232). Apart from the fact that quantum theories have great explanatory

[69] As easy problems, Chalmers (2007, p. 225) suggests explaining the following phenomena: 1) the ability to discriminate, categorize, and react to environmental stimuli; 2) the integration of information by a cognitive system; 3) the reportability of mental states; 4) the ability of a system to access its own internal states; 5) the focus of attention; 6) the deliberate control of behavior; 7) the difference between wakefulness and sleep.

advantages, such as its capacity to explain nondeterministic and nonlocal phenomena, and considering that it might seem natural to assume that quantum processes are involved in brain complexity, those theories too, claims Chalmers, tend to leave the question unanswered.

On Chalmers's view, the standard approach, including quantum theories, will not do: "to explain experience, we need a new approach. The usual explanatory methods of cognitive science and neuroscience do not suffice" (2007, p. 228). But why? Because "we need to know more than *which* processes give rise to experience; we need an account of why and how. A full theory of consciousness must build an explanatory bridge" (Chalmers 2007, p. 231). The proliferation of consciousness studies —even if apparently to no avail—has been rather negligible in terms of progress regarding the hard problem, but not because of hollow arguments or due to the complexity of the *explanandum*, but rather because many such attempts have focused on explaining something else (for instance, physical processes of some sort from which purportedly consciousness somehow emerges), or worse because consciousness has been explained away by glibly deeming it a nonexistent. Chalmers proposes then to build an explanatory bridge between the hard facts of our physical comprehension of the universe and that apparent 'extra ingredient' which is brought about by phenomenal consciousness.

On this account, consciousness is no doubt an *explanandum* in much need of a suitable *explanans,* but the variety of theories doing the job suggests that in this terrain reigns much confusion. As an example of this variegated explanatory range, let us quote at length a summary of consciousness research at the beginning of the twenty-first century provided by Frith & Rees (2007):

> We know that life does not depend upon a vital essence, but we are still not sure about consciousness. Perhaps there is a vital essence that turns a zombie into a human. There are various proposals as to the nature of this vital essence. Eliminative materialists (e.g., Paul and Patricia Churchland) have concluded that consciousness is itself a vital essence and therefore does not really exist. For *functionalists*, following in the footsteps of La Mettrie, the vital essence is a computational algorithm of sufficient complexity. This can be instantiated in silicon just as well as in neurons. If a machine has the right kind of complexity it will be conscious. No new physical principles will be required to understand how it works. Others claim that some as yet undiscovered scientific process such as quantum entanglement at a microscopic level, is needed to explain consciousness. And finally *mysterians* think that the problem of consciousness is so complex that the human brain can never explain it. (p. 18)

Consciousness, so it seems, is not amenable to experimental study, mathematical modelling, and the mechanistic worldview described by physicalism. Reductionist programs fail, as Nagel (1974) famously said, because one cannot reduce what one does not understand: "without some idea, therefore, of what

the subjective character of experience is, we cannot know what is required of a physicalist theory" (p. 437). No wonder consciousness has been more often than not plainly glossed over as a scientific topic, postponed as the later product of a series of mechanisms, or simply ridiculed as a metaphysical *élan vital*. So physicalists and mechanistic functionalists alike take consciousness to be a thing amongst other things—an *object*, for what is worth—residing in the human brain, or an emergent functional property resulting from unconscious mechanical processes due to computational organization, which can be integrated as part of natural reality, and then go on to puzzle over why it does not seem to rightly fit in, why it does not reside comfortably within this theoretical framework. The framework is, of course, never put into question, but instead consciousness is declared a mystery, and the brain, since a suitable theory is lacking, the most complicated system in the entire universe. On this account, however, Nagel's demand should not be ignored: "if physicalism is to be defended, the phenomenological features must themselves be given a physical account" (1974, p. 437). But as Fodor once said: "nobody has the slightest idea how anything material could be conscious" (1992, p. 5). So the challenge is to give an appropriate (that is, scientific) account of why this turns out to be the case and what it implies for the theoretical framework from which the problem is being approached.

Rejecting naturalism altogether and thus declaring consciousness out of the reach of the natural world would put us at the risk of throwing out the baby with the bathwater. So naturalists have argued that perhaps the picture is simply not complete, as though something were still missing to assemble the pieces of the puzzle. What is lacking in order to address the problem of consciousness is perhaps a theoretical breakthrough capable of bridging the gap between the biological workings of the brain—which ultimately depend on the laws of physics—and the phenomenal aspects accompanying those natural events—whose binding 'laws,' as it were, are yet to be uncovered. And more and more it is recognized that this breakthrough might be achieved with the introduction and dissemination of information theory.

Information theory provided the first crucial step for a mathematical account of cognition. Indeed, talk on the information revolution as modifying our perspective on the ultimate nature of reality and as providing consequently a change of metaphysical perspective, from a materialist one (in which physical objects play a key role), to an informational one, is nowadays ubiquitous (see Floridi

2010 and 2012). Information is currently widely regarded as being not only pervasive but conceptually all-encompassing: as the subtitle of Seife's *Decoding the Universe* (2007) would have it, the science of information "is explaining everything in the Cosmos, from our brains to black holes." Incorporating information theory into cognitive science helps in avoiding the rather eerie character of the mental, and it might count as a crucial step forward in solving the problem of bridging those scientific gaps mentioned by Kilian (1970) forty years ago: namely, the gap between the mental and the physical. It is expected that this may finally begin to clarify why something material can be conscious.

This change of metaphysical perspective, from a physical to an informational one, owes its foundation to physicist John Archibald Wheeler's 'it from bit' doctrine, according to which 'it' (every particle, every field of force, even space-time continuum itself) derives its function, its meaning, and its very existence from the apparatus-elicited answers to yes-no questions, binary choices: bits (1992). The doctrine suggests that not only information is fundamental to the physics of the universe but, most importantly, that even physical properties and laws may derive from informational properties and laws. The informational paradigm—although in its first, cybernetic, stage—was also behind the project of conceiving of the neuron, not simply as the most basic anatomical unit of the central nervous system, but as the basic information processing unit (McCulloch & Pitts 1943).

Let us take a closer look at how the informational paradigm can be incorporated into cognitive science research. In the most speculative chapter of Chalmers's *The Conscious Mind* (1996), this is precisely what is attempted: the inclusion of the most promising aspects offered by information theory in a fundamental theory of consciousness. On Chalmers's view, information is a nice candidate as a link for bridging the gap between the physical and the phenomenal, specially due to the double-aspect principle, which states that "whenever we find an information space realized phenomenally, we find the same information space realized physically. And when an experience realizes an information state, the same information state is realized in the experience's physical substrate" (1996, p. 284). The visual experience of color exemplifies how this double-aspect principle is realized. Indeed, the experience, argues Chalmers, is both lived phenomenally and realized physically in a neurally coded representation in the visual cortex. On Chalmers's account, the principle can be both generalized to other more

complex experiences and to physical reality as a whole, that is, this double realization of information could be fleshed out into a system of basic laws connecting the physical and the phenomenal domains:

> We might put this by suggesting as a basic principle that information (in the actual world) has two aspects, a physical and a phenomenal aspect. Wherever there is a phenomenal state, it realizes an information state, an information state that is also realized in the cognitive system of the brain. Conversely, for at least some physically realized information spaces, whenever an information state in that space is realized physically, it is also realized phenomenally. (1996, p. 286)

This is Chalmers's controversial hypothesis: that information is the key to the link between physical processes and conscious experience. However, what makes it controversial is not merely the incorporation of information theory into consciousness research, but its generalization to all physical reality. This assumption leads Chalmers to speculate that experience might be ubiquitous (an idea he himself concedes might be regarded as outrageous or even crazy).[70] The ubiquitous character of experience, on this account, implies the presence of experience in very simple systems. For all its strangeness, the idea of the ubiquitous character of experience—and let us remark: it is subjective experience what the parlance is here about—is at first sight not as far-fetched as it might sound. Certainly we can imagine, without committing flaws of reasoning, what it is like to be a bat, or what it is like to be a dolphin, or even we can suppose that there is something it is like to be a mouse. (Incidentally, it is said to be easier to imagine the what-it-is-like character of experience in creatures not as alien to us, like mammals, than it is in more alien forms of life, like insects.) Nagel has put it this way: "Conscious experience is a widespread phenomenon. It occurs at many levels of animal life, though we cannot be sure of its presence in the simpler organisms, and it is very difficult to say in general what provides evidence of it" (1974, p. 436). In spite of Nagel's caution regarding evidence as to how widespread the what-is-likeness character of experience in simpler organisms really is, Chalmers is certainly more willing to go further, so he asks "what is it like to be a thermostat?" (1996, p. 293). Now, it would have been adventurous enough to jump from mammals and imagine, for instance, what it is like to be a spider. But Chalmers is out to overthrow traditional ideas and blurs thus in one fell swoop the frontiers separating animality and mechanism.

---

[70] Someone who finds this idea crazy, says Chalmers, "at least owes us an account of *why* it is crazy" (1996, p. 295). However, it is certainly Chalmers the one indebted with such an explanation.

On Chalmers's speculation, the question is where consciousness should wink out as one moves along the scale from fish to slugs, through simple neural networks all the way to thermostats:

> the phenomenology of fish and slugs will likely not be primitive but relatively complex, reflecting the various distinctions they can make. Before phenomenology winks out altogether, we presumably will get to some sort of maximally simple phenomenology. It seems to me that the most natural place for this to occur is in a system with a corresponding simple 'perceptual psychology,' such as a thermostat. The thermostat seems to realize the sort of information processing in a fish or a slug stripped down to its simplest form, so perhaps it might also have the corresponding sort of phenomenology in its most stripped down form. It makes one or two relevant distinctions on which action depends; to me, at least, it does not seem unreasonable that there might be associated distinctions in experience. (1996, p. 295)

Not even the fact that a thermostat is not alive—perhaps the most obvious reason that would prohibit one to speculate about subjective experience being scattered all around in the universe: from stars to river pebbles—makes Chalmers backpedal a single step back, since in principle a disembodied silicon brain would arguably fail to qualify as alive, but it might be conscious (1996, p. 296). As it happens, mentioning river pebbles is not a rhetorical twist, for Chalmers has it that "one can find information states in a rock—when it expands and contracts, for example—or even in the different states of an electron. So if the unrestricted double-aspect principle is correct, there will be experience associated with a rock or an electron" (1996, p. 297). A rock, however, is not conscious in the same way a thermostat would be: "a rock, unlike a thermostat, is not picked out as an information-processing system. It is simple picked out as an object, so the connection to experience is less direct" (*idem*). Therefore, the generalization of the double-aspect principle and the enthusiastic endorsement of the 'it from bit' doctrine permits Chalmers to hold a species of panpsychism: understood not as the metaphysical foundation of his view, but as one way that the natural supervenience of experience on the psychical might work (1996, p. 299).[71] This is ultimately rooted in an *amplification* of information, which is crucial for Chalmers's view that information must be grounded in phenomenology: "experience is information from the inside; physics is information from the outside" (1996, p. 305). Chalmers feels thus compeled to offer a picture of the world as *pure* information, which rests on the assumption that the universe could be a giant computer (Lloyd 2007) and that all of space-time could

[71] The truth be told, Chalmers remains an agnostic of panpsychism, to which he merely refers as a possibility that can be taken into account, in spite of its counterintuitive character (see 1996, p. 299 and 1997, p. 165).

be grounded in computational processes (Vedral 2012). A picture he finds "strangely beautiful: a picture of the world as pure informational flux, without any further substance to it" (1996, p. 303).

It is not difficult at all to see why Chalmers's book caused such an unbridgeable divide between those saluting the book as one of the best contributions to the science of mind ever made, and those rejecting it as metaphysical speculation of the worst kind.[72] Among the critics, it is John Searle who has strongly rebuked Chalmers for introducing a metaphysical distinction between consciousness and physical reality. For Searle—a biological naturalist who thinks consciousness is simply caused by brain processes—the so-called mystery of consciousness stems from our current ignorance of all the details surrounding the workings of the brain, and so "the sense of mystery will be removed when we understand the biology of consciousness with the same depth of understanding that we now understand the biology of life" (1997, p. 201). But Chalmers's claim regarding the nonphysical character of consciousness entails that even understanding perfectly how the brain works will not do: "to explain why and how brains support consciousness, an account of the brain alone is not enough; to bridge the gap, one needs to add independent 'bridging' laws" (1997, p. 164-165), which, of course, Chalmers has tried to suggest by resorting to information theory.

Searle seems to take Chalmers's species of dualism at face value and then puzzles over how it is even tenable conceiving of psychological terms such as 'pain,' 'belief,' etc., as having two completely independent meanings: "one where it refers to nonconscious functional processes, and one where it refers to states of consciousness" (1997, p. 168). Despite this, Chalmers's property dualism entails that the physical structure of the world is logically consistent with the absence of consciousness. So consciousness is a further fact about the world which cannot be reduced to physical processes. Chalmers's claim is not that a painful experience produces two separate entities: the physiological reaction and the phenomenal experience. But rather that by simply explaining the performance of cognitive functions, the fact that these merely physical processes lead to experience remains unexplained. On Chalmers's account, information is the glue coupling physical processes with phenomenal qualities.

---

[72] The book has been widely reviewed by prominent figures such as Roger Penrose, Colin McGinn, David Papineau, Max Velmans and William Clancey. For a complete list of reviews, see the next link in Chalmers's website: <http://consc.net/books/tcm/reviews.html>.

A biological naturalist like Searle takes stock on the debate between dualists and reductive materialists by insisting that "one way to see biological naturalism is as an attempt to preserve what is true in each [e.g., in dualism and reductive materialism] while discarding the false" (2007, p. 333). Chalmers has advocated a nonreductive functionalism which might be regarded as a combination of functionalism and property dualism (1996, p. 249), and this strikes Searle as a *reductio ad absurdum* of the combination of the two doctrines, in that Chalmers thinks "you can keep your functionalism but you should add property dualism to it. The result, in my view, is to trade one false doctrine for two" (1997, p. 163). Conversely, the position that biological naturalism defends is that the powers of subjective consciousness are perfectly consistent with their causal functioning as a natural neurobiological set of processes (Searle 2007, p. 331). Consciousness is at the same time qualitative, subjective, and unitary, and so any attempt to reduce consciousness to something else and thus losing the aforementioned characteristics is due more to traditional philosophical disputes (the Cartesian 'mistakes,' for example) which have become common sense to the philosophers, than to the purported mysterious aura surrounding consciousness.

The last Searlean objection to Chalmers, has to do with Chalmers's assumption that information is pervasive and so, in a certain way, the counterargument can be applied also to the pervasiveness of computation in the universe. Searle's famous *Gedankenexperiment* of the Chinese Room (1980) was meant to demonstrate that semantics is not intrinsic to syntax and thus it presented a challenge to the claim that it was possible for computers running a program to have a mind, to understand, and even to have consciousness. Given that computation is intrinsically syntactic—for a computer simply manipulates formal features—understanding cannot arise from computation. Searle modified later the scope of his argument to show that, on a closer inspection, he even bit the bullet with Strong AI conceding too much to it (in that he argued the theory was at least false), whereas it must be granted that it is incoherent (1997, p. 14). And this is why: "computation is not an intrinsic process in nature like digestion or photosynthesis, but exists only relative to some agent who gives a computational interpretation to the physics. The upshot of Searle's counterargument is that computation is not intrinsic to nature but is relative to the observer or user" (1997, p. 14-15). Of course, anything can be interpreted computationally—a window, for instance, can be viewed as a computer consisting of a two-

state instruction manual: 1=open and 0=close—but, on Searle's account, that does not make the terms 'symbol,' 'syntax,' and 'computation' intrinsic features of nature like 'tectonic plate,' 'electron,' or 'consciousness.' This is why Searle now thinks that this is a different argument than the one offered with the Chinese Room. It is slightly different but it is also deeper because it shows, not only what the former already showed (that semantics is not intrinsic to syntax), but above all that syntax is not intrinsic to physics. Syntax is too an assigned, human phenomenon. By the same token, information is not intrinsic to the physical constitution of reality. It can be granted that rain hitting the ground contains information because it makes changes in the ground, but this, for Searle, is observer-relative, as are, say, the concentric rings of a cut tree which are used to calculate its age.

However, the discussion between Chalmers and Searle does not ensure on its own that we have gotten rid of the hard problem. In view of the wide ramifications of the topic, the exposition of some further theoretical problems of phenomenal consciousness is forthwith needed.

Theoretical Problems of Phenomenal Consciousness

Now it is time to sort out the possibility of carrying out a reduction of consciousness to physical, even molecular, mechanisms, or to discard it and proclaim conversely its irreducibility. The latter conclusion amounts to deeming the hard problem intractable. The reason for this apparently intractable character of consciousness needs further exploration. At least on occasion the attempt is also made in order to render the concept of consciousness otiose.[73] But first, of course, it is all the more important to have a clear idea about that which is supposed to be reduced or about that which one is supposed to gloss over altogether. As a matter of fact, relinquishing some clarity on this regard leads frequently to conceptual muddles. Hence, it seems prudent to explore the scope of the concept of phenomenal consciousness, which has unexpectedly taken on renewed currency after decades of denial and being treated with patronizing contempt.

Let us remember that the gist of Nagel's argument regarding the irreducibility of consciousness (the what-is-likeness of experience) runs as follows: unlike the facts of physical reality, which are

[73] This will be the topic of the next chapter.

objective and are thus accesible from a public and generalizable perspective, "every subjective phenomenon is essentially connected with a single point of view, and it seems inevitable that an objective, physical theory will abandon that point of view" (1974, p. 437). Therefore the facts of phenomenal experience are by definition different to the facts of physical objective reality: "for if the facts of experience—facts about what it is like *for* the experiencing organism—are accesible only from one point of view, then it is a mystery how the true character of experiences could be revealed in the physical operation of that organism" (1974, p. 442). The hard problem of consciousness construed as the what-is-likeness of experience is here more recalcitrant than ever, for it is possible to possess all the neurological facts about the human brain, just as it is possible to be quite knowledgeable about bat neurophysiology, and that without shedding a single shred of light on what it is like to be a bat or, for that matter, on what it is like to be us. For Nagel, a reduction of the first-person perspective characterizing subjective experience to an objective third-person perspective is as absurd as though one wished to understand phenomenal consciousness by means of abandoning what is most intrinsic and particular of it (1974, p. 444). It would be as preposterous as pretending to know what it is like to be us by resorting to an objective perspective about human beings that could be available to extraterrestrial scientists studying human neurophysiology, who definitely cannot know what it is like to be us without our human point of view, just as we cannot really know what it is like to be a bat by merely resorting to a corpus of knowledge which we already possess, namely bat neurophysiology, since the bat point of view—for all our acumen and knowledge of the neurophysiological facts—would be utterly missing. On Nagel's view, greater objectivity means less attachment to a specific viewpoint, and if precisely it is this specificity what characterizes subjectivity, a greater objectivity "does not take us nearer to the real nature of the phenomenon: it takes us farther away from it" (1974, p. 445).

So a 'ruthlessly reductive account,' like the one intended by Bickle (2003), will not do. In Blickle's rather radical view, what must be sought is strictly the lowest possible level of explanation: the *mind-to-molecule reduction*. Like all types of reductionism, the one espoused by Bickle intends to strip consciousness (which is 'nothing special,' on his account, see 2003, p. 189) of its air of mysticism, because it is nothing but a pet word so characteristic of armchair metaphysicians. Accordingly, at least for the time being, higher-level explanations, like the ones one is to find in cognitive psychology and

cognitive science, may be tolerated as long as no lower-level (that is, molecular) explanations are on offer (2003, p. 115). Bickle and Ellis (2005) quote extensively from a study (Lui & Newsome 2000) that claims that microstimulation experiments on awake, behaving monkeys have established causality between activity of specialized cortical neurons and controlled behavior. Drawing from theses results, Bickle and Ellis feel free to conclude that phenomenological experiences are induced by neural events, rather than external stimuli, and so determine the entire perceptual experience of a subject (Bickle & Ellis 2000, p. 140). This is the so-called 'single-cell approach:' all of our perception of the external world as well as our internal thoughts can be said to result from patterned electrical activity among the several billion neurons that comprise the central nervous system.[74]

It bears emphasizing that—despite claims to the contrary—there are several important objections, not to the neurophysiological fact that neural events are involved in perceptive experience, but to the overall consequences of this reductive approach for philosophy and the so-called higher-level sciences. In Bickle's expectations, for instance, higher-level scientific discourses, such as the ones to be found in cognitive psychology and cognitive science (don't even mention philosophy!), are bound to disappear (2003, p. 3-4) and little by little will yield place to a 'molecular philosophy.' In spite of this, there is ample evidence that an ecological approach—in which interactions between the organism and its environment are crucial—is needed for a suitable understanding of the whole organism. In fact, understanding an organism requires a focus on the entire agent-environment system, and not only exclusively on system-internal properties (see Looren de Jong & Schouten 2005, p. 481). The whole story then cannot be merely reduced to cortical microstimulation.

There is finally another reason for suspicion. One is justified to be dismissive towards such project of construing the role of philosophy as mere neuroscience interpretation and extrapolation thereof, given that the the task of philosophy of science is to concern itself with the foundations, methods and implications of the scientific enterprise. In this vein, Bickle remains in debt as a

[74] Strangely enough, Bickle and Ellis think Husserl's phenomenological reduction and some of Merleau-Ponty's ideas bear somehow resemblance with their species of mind-to-molecule reduction: "What is certainly clear is that both Husserlian phenomenology and Merleau-Ponty's attempts to mix it with psychology and neurophysiology are quite consistent with such odd findings as cortical microstimulation of conscious perceptual states" (2005, p. 160).

philosopher of science[75] because it is certainly dubious to assume without further justification that the criteria for a successful explanation of cognition are decided by means of explicating simpler processes. What invisible power entitles Bickle to mandate that this assumption must not be discussed? It is only on behest of prejudice that philosophical questioning is deactivated. And philosophical questioning is *par excellence* critique of prejudice. The scope of the micro-macro relation is no doubt in need of philosophical treatment, for it cannot merely be assumed. However, this relation, that puts the macro at the mercy of the micro, remains largely unexplored by Bickle. There is indeed a panoply of central philosophical questions which cannot be answered from the perspective of the molecular expert. It may even be true that philosophy and the natural sciences do not just have a different perspective on the same topics, but different topics altogether (see Lembeck 2010, p. 176). And, again, assuming this difference, or conversely, taking for granted an identity between science and philosophy, is certainly a discussion of exogenous nature to the practice of the expert. Conflating unabashedly philosophical questioning with justified empirical questions as the ones related with molecular processes (a conflation which stands as the chief claim of scientism), threatens to cloud both our understanding of science and philosophy.

With these caveats noted, the question still remains: even if objective processes are spotted in neural activity that are directly related with phenomenal consciousness, does this (third-person) objective (and molecular) perspective give us a glimpse on the hard problem? Does that eradicate it or render it otiose, as Bickle claims?

That the problem remains refractory to philosophical closure can be exemplified by resorting to the case of qualia, taken as qualitative properties accompanying subjective experience which are claimed to be nonreducible. Qualia is plural for *quale* (from the latin 'of such kind') and, albeit introduced in philosophy by Charles Sanders Peirce, it was C. I. Lewis who extended its customary use in the philosophy of mind:

> Qualia are subjective; they have no names in ordinary discourse but are indicated by some circumlocution as 'looks like;' they are ineffable, since they might be different in two minds with no possibility of discovering that fact and no necessary inconvenience to our knowledge of objects and their properties. All that can be done to designate a quale is, so to speak, to locate it

[75] Looren de Jong and Schouten consider his treatment of the implications for philosophy of his mind-to-molecule reduction "philosophically shallow," 2005, p. 482.

> in experience, that is, to designate the conditions of its recurrence or other relations to it. Such location does not touch the quale itself; if one such could be lifted out of the network of its relations, in the total experience of the individual, and replaced by another, no social interest or interest of action would be affected by such a substitution. What is essential for understanding and communication is not the quale as such but the pattern of its stable relations in experience which is what is implicitly predicted when it is taken as the sign of an objective property. (Lewis 1929, p. 124-125; quoted by Bayne & Montague 2011, p. 11)

As Bayne and Montague have argued, "we are presented here with a dichotomy between what is experiential and non-conceptual on the one hand, and what is conceptual and non-experiential on the other hand" (*idem*). No wonder the discussion surrounding qualia is plagued with controversy (with *qualiaphiles* and *qualiaphobes* defending wholeheartedly their viewpoint),[76] partly as a result of C. I. Lewis's entrenched idea that qualia must be conceived of as nonconceptual features of experience. Lewis's view that the designation of the quale does not even touch it, has been taken as meaning that qualia must be then ineffable. As a matter of fact, for qualiaphobes, qualiaphilia can be seen as a a worrisome sign having undesirable consequences for objectivity, since the case for 'indirect realism' mounted by qualia defenders implies an indirect access to reality that runs the risk of proliferating skepticism and an objectivity gap. And this as if the real were hidden in a veil of sensation. Qualiaphiles can be thus easily associated with relativism and solipsism. According to Wright, "both the righteously minded among philosophers and the tough-minded physicalists are tempted to reject the proposal [that the introduction of qualia into philosophy means worth the effort] out of hand, the former for its rendering eternal verities dubious, particularly truth and objectivity, and the latter for its determinedly trying to smuggle occult entities into science" (2011, p. 341). The view of qualia critics is that turning to nonconceptual experiencing creates more problems than it solves. One need only remember McDowell's warning against the assumption accompanying the endorsement of nonconceptual content that there can be somehow bare sensing: beware of the Myth of the Given as a brute effect of the world! (2000, p. 42).

However, it must be stressed that for qualiaphiles this completely misses the point. There are ways of skewing the arguments mooted by qualia defenders: for instance, mocking them as though they were still arguing for 'red neurons in the brain,' consequently threatening "to make experience itself

[76] See arguments for and against qualia in the collection of essays edited by Wright (2008). For related discussions regarding phenomenal consciousness by appeal to sensory experience, see the collections edited by Heckmann & Walter (2006), Pauen, Schütte & Staudachter (2007) and Bayne & Montague (2011). Other sources worth noticing are Metzinger ed. (2006) and Michel (2011).

something mysterious and inaccesible" (O'Regan & Noë 2001, p. 962). According to Wright, who has presented a laborious (and very convincing) account of what it means to believe in qualia, this objection constitutes the typical way of trying to set down both a pictorial interpretation of qualia (deriving from the old sense-datum theory) and qualia theory "in the chronicle of philosophical defeats" (2011, p. 342). But let us take a closer look at the qualia-approach considered as indirect realism; an approach which accordingly loathes the transparency-case of direct realism and objectivism.

Foisting on indirect realism the idea that qualia are objectionable because of their being regarded as "objects of an internal gaze" (Evans 1982, p. 231), is potential for confusion. Conversely, for qualia advocates, qualia are strictly speaking no objects at all. Qualia are nonepistemic, in the sense that sensory experiences can be said to carry no information whatsoever about entities, but are merely evidence which can be construed according to the motivations of the observer (Wright 2011, "Introduction," p. 4). Sensations are 'semantically inert' in that they do not convey information that was already there but function only as evidence for further perceptive dealings. Nonetheless, enthusiasts of the 'it from bit' doctrine have it that information is pervasive in the universe. According to a strange view espoused by Tye, for instance, "before any human noticed rings inside trees, the number of rings represented the age of the tree, just as it does now" (1997, p. 100). This invites the obvious retort: represented for whom? Are not such concepts as 'age' and 'tree' human designations? Are they not only meaningful in the human referential context? Such an assertion appears indeed wholly unintelligible for the indirect realist: there was literally not any information in the number of tree rings—to say it with Searle: here too information is not simply discovered in the fabric of physical reality, but it is assigned; one does not discover information in nature, but renders an informational interpretation of it. On the opposite side, Evans (1982) characterized the nonconceptual precisely as that information which was there, awaiting to be picked up and conceptualized: "the informational states which a subject acquires through perception are *non-conceptual*, or *non-conceptualized*" (p. 227). On the indirect realist's point of view, however, conceiving of the tree rings as evidence renders the case nonepistemic. According to Wright, "whenever Evans speaks of 'nonconceptual content' it is always in terms of discrete singular entities awaiting perception. 'Information' about entities is already there in the sensations" (2011,

"Introduction," p. 21). The indirect realist or qualia advocate denies this. Therefore, Evan's (and, for that matter, McDowell's) 'nonconceptual' cannot be simply equated with 'nonepistemic.'

This is a very important point for the indirect realist to make, since "the non-epistemic precedes the epistemic" (Wright 2011, p. 348). And making this point permits Wright to refute the arguments that qualia are somehow ineffable and incorrigible. The assumption that our encounter with reality is at first nonepistemic and that evidence precedes conceptuality, implies precisely that I can be corrected about that evidence. I can be corrected about the accuracy of my judgement. That first evidence which I encounter when I see a bird taking off from a tree and that my friend makes me reconsider, for it was actually two birds and not only one, shows, first, that what I saw—albeit mistakenly—is not in the least ineffable, and, second, that the nonepistemic allows for corrigibility. Should the contrary be correct, then it would not be possible for me to be updated in my knowledge by someone else who 'saw it better.' This is why learning in perception is possible: "the very fact that one person can alter another's percepts, whether or not that alteration is successful, is a proof of the existence of a non-epistemic nature of the sensory fields of both observers, since it allows a play of percepts" (Wright 2011, p. 352). But doesn't this imply a resurrection of the Myth of the Given? Not at all. The point is precisely that evidence is hermeneutical for it necessitates interpretation: "human interest is manifestly relevant in determining for what purposes the evidence shall be interpreted" (Wright, 'Introduction,' p. 4). So Tye's assertion as to information being already contained in the tree rings is unintelligible. There is no such thing as bare information which somehow couples miraculously with our concepts of it. On the contrary, *our concepts are also interpretations*. Therefore, Tye's argument can be confronted by two questions: who is selecting what parts of the rings count as rings? Moreover, what is to count as a year? The always interpretable structure of our concepts, prohibits that we forget that they are human.

Wright is convinced that conceiving of subjective experience as a nonepistemic field comports the important evolutionary advantage that the object appears as an interpretable selection that remains alterable (2011, p. 347). The objectivist wants to begin accounting for our encounter with the real equating it with an encounter with objects, whereby single material objects are said to precede the perception of them (Matthen 2005). Conversely, the indirect realist argues that to sense is not to know or that sensing is not perception. Sense perception is thus a misnomer. Assuming the precedence of

concepts distorts the evidence field, for it introduces as given what is merely the interpretable and alterable play of concepts. On Wright's view, conflating the relation between sensing and perceiving otherwise, amounts to the impossible perfect union of word and world (2011, p. 357). So, to use a distinction about consciousness introduced by Block (1995), phenomenal consciousness (the sort of consciousness when we see, hear, smell and taste) precedes access-consciousness (the sort of consciousness when we talk and think).

According to the aforementioned points, the case for qualia can certainly not be overthrown by simply espousing a nay-saying attitude neglecting thereby the possibility that there are these special qualitative raw feels accompanying conscious experience. What Wright's comprehensive treatment of the topic must have made clear is that qualia are not to be confused with entities of any kind. And it is only because qualia have been so construed, as occult entities, that the critics tend to fear that experience be relegated to the inaccesible and mysterious. On the deep philosophical side of the issue, Wright thinks that "*objectivity* can never be equated with *existence*" (2011, p. 352), which is, incidentally, a false equation assumed by those who do not consider the evidence field and its nonepistemic dimension.

However, we may indeed have succumbed to the illusion of the real. And one should not want to do as Descartes who doubted everything that could be doubted, except for his qualitative, what-it-is-like subjectivity (see Dennett 1993a, p. 381). Accordingly, the next chapter explores the possibility that consciousness might be just an illusion.

| 7 |

THE REALLY ASTONISHING HYPOTHESIS

Consciousness as Illusion

Daniel Dennett has claimed for many years that phenomenal consciousness is not that special.[77] The problem is mainly that the common parlance of subjective raw feels and all the vocabulary associated with the problems of phenomenal consciousness appear highly speculative and belong more to armchair philosophizing than to a scientific research program. Hence qualia, according to Dennett's legendary paper (1993a), must be *quined*, that is, their existence must be resolutely denied. So instead of admitting in our theoretical imagination of speculative forays wandering towards a nonepistemic evidence field, philosophy should take a look at how things are done in cognitive science. On Dennett's assessment, within cognitive science the method of heterophenomenology has been utilized to render subjectivity as something less mysterious and occult. The method, admits Dennett, is not his own creation. His only task has been to trot it out in the very practice of cognitive science:

> Heterophenomenology is nothing other than the scientific method applied to the phenomena of consciousness, and thus the way to *save* the rich phenomenology of consciousness for scientific study. I didn't invent the heterophenomenological method; I just codified, more self-consciously and carefully than before, the ground rules already tacitly endorsed by the leading researchers. (Dennett 1993b, p. 50)

In his self-presentation for the entry 'Daniel Dennett' in *A Companion to the Philosophy of Mind* (Dennett 2005, pp. 236-244), he traces his own thinking path spanning from his early attempt (Dennett 1969) at a theory of content, that is, of intentionality, on top of which a suitable theory of consciousness could be build. According to his own appraisal, this theory only found a satisfactory form in his opus magnum, *Consciousness Explained* (1991), which nonetheless presupposed the theory of content presented in *The Intentional Stance* (1987). Dennett admits that "from the outset I worked from the 'third-person point of view' of science, and took my task to be building—or rather sketching the outlines of—a physical structure that could be seen to accomplish the puzzling legerdemain of the mind" (2005, pp. 236-237).

---

[77] Indeed, he affirms jokingly that "consciousness is not that supercalifragilistiexpialidocious as many people like to believe" (2006, p. 42).

Dennett's theory of intentionality, along with his well-known 'metaphysical minimalism' (Dennett 1982, p. 159; see Zahavi 2007, p. 37), have gained him the fame of being a heterodox theorist of mind. And this—for Dennett's despair—given that many philosophers traditionally have an initial allegiance to the realm of the mental and to the deliverances of introspection (Dennett 2005, p. 237).[78] Dennett's radical treatment of folk psychology according to which people's most basic beliefs are to be equated with nothing but fiction—since beliefs should be treated indeed as theorist's fictions (Dennett 2003, p. 20; a position defended also in Dennett 1971, 1987 and 1991)—indicates his conviction that the world, and our whole experience thereof, is nothing but a grand illusion. The world *seems*, no doubt, as meaningful, as being loaded with all kinds of events which we experience subjectively. However, following Dennett's absolute cognitivism, what it is like to have those experiences might just be the resulting by-product of a series of blind mechanisms. There is no way to be sure about the status of those phenomenal events, at least until a unifying theory is found that rightly explains the matching mechanisms between those meaningful events of experience with the 'real goings-on' in the brain. It must be reminded that there is a 'dirty secret' (Searle 1997, p. 198) in contemporary neuroscience: so far there is no unifying theoretical principle of neuroscience in the way there is an atomic theory of matter, a germ theory of disease, a genetic theory of inheritance, a tectonic plate theory of geology, a natural selection theory of evolution, etc. Therefore everything is open until the personal level of experience has been matched with subpersonal physical mechanisms. Only then, that is, when the gap between the personal and subpersonal levels be scientifically surmounted, it will consequently be possible to declare that the experiences comprising subjectivity are more than just autophenomenological beliefs or plain rubbish. On Dennett's view,

> if we were to find real goings-on in people's brain that had enough of the 'defining' properties of the items that populate their heterophenomenological worlds, we could reasonably propose that we had discovered what they were *really* talking about… And if we discovered that the real goings-on bore only a minor resemblance to the heterophenomenological items, we could reasonably declare that people were just mistaken in the beliefs they expressed. (Dennett 1991, p. 85)

Whether there are neural correlates of consciousness from whence a suitable unifying scientific theory will emerge, it is not for philosophy to decide. That is entirely an empirical question. But, according to

[78] "My insistence on the need for philosophers to stoke up on the relevant science before holding forth, and my refusal to conduct my investigations by the traditional method of definition and formal argument, have made me a distinctly impure philosopher of mind" (2005, p. 243).

Dennett, what philosophy can certainly not afford is to speculate without finding support in science. So he has defined his own (philosophical) work as advocating a 'mild realism' that derives its explanatory power from the perspective of evolutionary models and reverse engineering (2005, p. 242). What is important is giving an account of the vices to which philosophers (more often than not) succumb; errors that derive their existence from the power of fictionalization entangled in the psychology of the folk. The question is not then what conscious experience is but rather how such fictive capacity for qualitative experiencing and the consequent expression about phenomenal events has originated in a complex intentional system composed of billions of microscopic robots. In fact, Dennett's philosophical precautions—his inclination developed at Oxford as a graduate student to distrust the methods he saw other philosophers employing (see Dennett 2005)—are due to his suspicion that folk-psychological views on experience cannot actually meet the standards and requirements of a scientific research program.

However, this last resort against folk psychological views does not amount to denying that dimension. Dennett does not neglect that there is a 'what-it-is-like' level of experience. His argument is just that those seemings that populate the level of what it is like to be us, human agents, are to be treated "as denizens of a theoretical fiction, characters in the subject's autobiographical novel, the default position of heterophenomenology until we do science" (Dennett 2007, p. 262). There is a distinction to be made between folk craft and folk theory: it is the making of a theory out of human folk craft what constitutes an ideology. Accordingly, "the *theory* of folk psychology is the ideology about the craft" (1998, p. 82). This last point has been couched by Dennett in terms of comparing folk psychology with folk physics. Folk physics (that objects can be stiff, that a ball will bounce when thrown over the ground, etc.) is true in the main, but the details of why those wonderful things are possible is the responsibility of physical theory. That, however, does not compel one to want to eliminate folk physics. The same can be said of folk psychology: it might be true in the main and one should expect it to be "similarly rich in retrospectively confirmed truth. But we can't count on it" (2006, p. 35).

So the fiction arising from folk psychological views is meaning-giving and it might work for us, as we know it does, but if one adopts the evolutionary perspective advanced in *The Intentional Stance*, it

must be granted that higher-level conscious processes can be decomposed "into hierarchically structured teams of ever more stupid intentional systems" (Dennett 2005, p. 240). One needs reminding that this argument is different than the one implied in Chalmers's question 'what it is like to be a thermostat?' Whereas thermostats are definitely intentional systems in Dennett's terms, Dennett is not willing to pose the question that Chalmers found so puzzling: at what point going down the scale of intentional systems does consciousness wink out? This question is one Dennett prohibits to ask for it is meaningless (it is only posed because ever since the dawn of philosophy, philosophers have been striving to find a 'prime mover,' a primitive something from which everything else can be said to nicely derive). Therefore questions regarding the point in the diminution of consciousness as one descends to simple neurons, or at what point in evolutionary history one is entitled to speak of genuine function, genuine selection-for, and not mere fortuitous self-preservation of entities endowed with a replicative capacity; all those questions encounter in Dennett the same demand: don't ask. Dennett puts it bluntly: "many of the most interesting and important features of our world have emerged, gradually, from a world that initially lacked them—function, intentionality, consciousness, morality, value—and it is a fool's errand to try to identify a first or most-simple instance of the 'real' thing" (2005, p. 240). From this follows that quining qualia *à la* Dennett is tantamount to stripping the hard problem of consciousness of its air of mystery in one fell swoop. There is, on Dennett's view, no hard problem of consciousness. But there seems to be one, there seems to be a mystery that makes one to go on and puzzle over why subjective experience seems so rich, vivid, and personal.

These assumptions let Dennett to present his theory of consciousness, the Multiple Drafts model, as opposed to the traditional Cartesian Theatre, which supposes the mind encounters something given and responds to it with interpretative judgements. In the traditional (Cartesian) view, a raw given of stimulation is supposed, which is then processed is some way and sent back to headquarters. In the Cartesian model, discriminations in all modalities are somehow put into registration and 'presented' for subjective judgment (Dennett & Kinsbourne 1992, p. 183). The Multiple Drafts model procedes otherwise and inverts this picture, since there is no place where it all comes together and no definitive crossing line dividing preconscious processing and the beginning of conscious appreciation:

> the alternative, Multiple Drafts model holds that whereas the brain events that discriminate various perceptual contents are distributed in both space and time in the brain, and whereas the temporal properties of these various events are determinate, none of these temporal properties determine subjective order, since there is no single, constitutive 'stream of consciousness' but rather a parallel stream of conflicting and continuously revised contents. (*idem*)

So if philosophers thrust themselves into the search of the place and the time of consciousness, they will find nothing but mechanisms. The magic of consciousness will lose its appeal just like when a magic trick is explained. Incidentally, this is exactly the purpose of Dennett's (1991) *opus magnum*: *Consciousness Explained* is an invitation to explain all the magic tricks surrounding consciousness. Stripping consciousness of its mysterious character is Dennett's debunking philosophical enterprise.

But it must be remarked that, for Dennett, heterophenomenology is not yet a complete science of mind but a humble previous stance to that future science—although it certainly hints at it. Its ultimate goal is to get things straight from the point of view of science and not to speculate and smuggle into the scientific discourse assumptions of a latter-day vitalism (Dennett 2005, p. 238). The resort to fiction is only momentary until a sound solution is found regarding "the problem of spanning the various explanatory gaps between the (first-)personal level and the subpersonal level of the natural sciences" (Dennett 2007, p. 268); a problem "about as difficult… as science—or philosophy—has ever faced" (*idem*). This is the real problem a theory must meet. The so-called hard problem, on the contrary, might be relegated to the level of the pseudoproblems and 'category mistakes' Dennett finds everywhere parasitic in the thoughts of mind theorists, which need a good casting-out.[79]

Now, a thorough explanation of heterophenomenology requires first and foremost the recognition of its negative character, which can only be clarified along with its target of criticism (a target that, on Dennett's view, exemplifies the iron presence of the Cartesian mind-set): autophenomenology. Autophenomenology can be designated, with Soldati, as 'naïve phenomenology,' that is to say, it consists in foisting on the folk craft real theoretical powers. Autophenomenology is then not a suitable theory—not even the beginning of one—but rather ideology in that it pretends, without success, to make a theory out of the fictive craft of folk psychological views.

---

[79] Dennett thinks he has concentrated on substantial (not merely on toy) problems such as the frame problem, problems about mental phenomena and 'filling in,' the binding problem and the problem of temporal anomalies: "I take these to be the real, as opposed to artefactual, problems of mental representation" (2005, p. 243).

On a terminological note, it must be granted that Dennett tends to be somewhat ambiguous with his definition of phenomenology. On the one hand, in some passages he simply denies that there is such a thing as phenomenology. For example, in the following: "while there are zoologists, there really are no phenomenologists: uncontroversial experts on the nature of the things that swim in the stream of consciousness" (1991, pp. 44-45). But on the other hand, heterophenomenology appears sometimes as the salvation of the phenomenological tradition and its problems. This salvation implies, in Dennett's words, to strip the phenomenological tradition of the 'anti-naturalistic ideology' that has somehow weighed it down in order to "salvage all the good ideas of Phenomenology and incorporate them into heterophenomenology" (Dennett 2007, p. 267).

This rather concessionary attitude towards phenomenological philosophy (that it has 'good ideas') is perhaps due to the occasion, since "Heterophenomenology Reconsidered" (2007), is a reply to a series of essays (edited by Alva Noë as a collective issue for the journal *Phenomenology and the Cognitive Sciences*), some of which were written by philosophers working in the phenomenological tradition who pointed out openly both Dennett's several misrepresentations of phenomenology and his notorious lack of familiarity with the object of his criticism, that is, the phenomenological tradition. Dennett has replied, that he obtained a certain familiarity with Husserlian phenomenology first as an undergraduate at Harvard with Føllesdal and then with his doctoral supervisor at Oxford, Gilbert Ryle, whom he credits as a masterful *connoisseur* of phenomenology (Dennett 1994). In his response to Carr, a Husserl scholar who has translated *Philosophie als strenge Wissenschaft* into English and who has been criticizing Dennett's rather *sui generis* concept of phenomenological philosophy, he has asserted that "part of what I thought I learned from those early encounters is that reading the self-styled Husserlians was largely a waste of time; they were deeply into obscurantism for its own sake" (1994). Although Dennett has admitted he is not strictly speaking scholarly acquainted with the phenomenological tradition nor has kept up with the relevant literature ("if I can figure out at least most of it without having to subject myself to all that stuff, why should I bother raking through it for further good bits? Life is short"), he has quickly disregarded any accusation as to whether he has been involved in the suspicious attitude of bad reading habits and prejudice: "it is precisely because my disregard has *not* been complete that it has been, and continues to be, so confident" (*idem*). This assertion might appear startling but it makes sense

if one considers that, for Dennett, phenomenology can be said to have at most refined the autophenomenological gaze of subjective seeming, but he is certainly not interested in the details.

Be that as it may, it must be noted that Dennett's capitalized use of 'Phenomenology' indicates that he distinguishes among various uses of the term. Among Dennett's writings, Zahavi (2007) has encountered at least three kinds of senses of 'phenomenology:' (1) Capitalized 'Phenomenology' refers to the phenomenological tradition founded by Husserl in Germany and continued by other European philosophers such as Heidegger, Fink, Sartre, Levinas, Merleau-Ponty, etc. (2) Non-capitalized 'phenomenology' refers to the notional seemings and experiential sensing of subjective character which are expressed in people's beliefs about their experience. This incidentally is a nonphenomenological appropriation of the term, because with it Dennett makes reference to a sort of ineffable private experience, which is Dennett's equation of phenomenology with qualia and the Nagelian 'what-it-is-like' level of experience (where putatively those seemings that are part of undergoing the very experiential level take place). This noncapitalized usage of the term is christened 'autophenomenology' by Dennett and it refers to the mental states by means of which subjects are capable of having a private experiential level of what-it-is-like for them to undergo those very experiences: experiences such as seeing colors, feeling pain and so on. Under autophenomenology, Dennett seems to include both aspects of consciousness differentiated by Block (1995): phenomenal consciousness (or P-consciousness) and access-consciousness. This last species of phenomenology so characterized by Dennett is, of course, the target of criticism of the third kind of phenomenology, that is, (3) Dennettian heterophenomenology.

However, Dennett's ambiguity remains and this distinction is not always respected. Along the way, phenomenology as defined in (1) and (2) seems to have been conflated and the early characterization of autophenomenology as some sort of "introspectionist bit of mental gymnastics" (Dennett 1987, p. 153) which is known in the psychological tradition of Wundt, for example, has been thoroughly attributed by Dennett to phenomenological philosophy. According to this, in a widely cited passage which is now worth quoting at length, Dennett characterizes Phenomenology (with a capital P) as the philosophical school or movement that

> grew up early in the twentieth century around the work of Edmund Husserl. Its aim was to find a new foundation for all philosophy (indeed, for all knowledge) based on a special technique of

> introspection, in which the outer world and all its implications and presuppositions were supposed to be 'bracketed' in a particular act of mind known as the *epoché*. The net result was an investigative state of mind in which the Phenomenologist was supposed to become acquainted with the pure objects of conscious experience, called *noemata*, untainted by the usual distortions and amendments of theory and practice. Like other attempts to strip away interpretation and reveal the basic facts of consciousness to rigorous observation, such as the Impressionist movement in the arts and the Introspectionist psychologies of Wundt, Titchener, and others, Phenomenology has failed to find a single, settled method that everyone could agree upon. (1991, p. 44)

It is widely accepted that introspection proved to be an unreliable source of scientific data. One needs simply reminding that Titchener and his disciples at Cornell University believed they had demonstrated that nonsensory conscious thought was impossible, whereas Külpe and the Würzburg School had demonstrated precisely the contrary to be the case.[80] In proclaiming the premature death of phenomenology, Dennett needs only to establish the link between phenomenology and introspection psychology. Phenomenology so construed, can be said to have been born already dead.

Apropos, Dennett is not alone in proclaiming phenomenology's conceptual inefficiency. In *Being No One* (2003), Thomas Metzinger has referred to phenomenology as an impossible theoretical endeavor: "you can take phenomenology seriously without running into all of its traditional problems" (2003, p. 591), which means: you can grant the existence of first-person experience, without resorting to phenomenological unsolved and largely obscure problems. What kind of problems is Metzinger talking about? In the same Dennettian heterophenomenological spirit, Metzinger goes on to affirm that "first-person access to the phenomenal content of one's own mental states does not fulfill the defining criteria for the concept of 'data'. My politically incorrect conclusion therefore is that first-person data do not exist" (*idem*). So there really is no such thing as a true phenomenological philosophy and this is why Metzinger has gone so far as to affirm the bankruptcy of phenomenology and its lack of scientific relevance: phenomenology is "a discredited research program… intellectually bankrupt for at least 50 years" (1997, p. 385).

So, for similar reasons, Dennett and Metzinger are both very dismissive of phenomenological philosophy and they certainly do not mind having a wrongheaded account of it. Even if they commit mistakes here and there when speaking with rather weak textual basis, that is, even in showing their lack of acquaintance with the relevant phenomenological literature, they think they are erring on the side of

[80] This anecdote is told by Lycan (1986), p. 21.

safety. For them, the appeal to first-person experience is a no starter and a dangerous move: "moving from the third-person to the first-person point of view is just asking for trouble; you get no data not already available to all of the rest of us from the third-person point of view, and you risk sending yourself off on wild goose chases trying to pin down conscious experiences that you only think you're having" (Dennett 2006, p. 49). No matter how fine-grained phenomenological descriptions of personal experience are, the heterophenomenologist, again, does not care about the details. This is the reason why waging a textual war, and merely referring phenomenological texts taking distance from the observation of subjective sensing, will not do.

In order to confront the claim that phenomenology belongs to the story of defeated philosophical movements, it would be more effective to give an answer to Dennett's own question, posed when discussing with Chalmers: "is there anything about experience that is not explorable by heterophenomenology? I'd like to know what" (2001). Dennett's challenge can be met by showing the constitution of experience, which is constantly neglected in Dennett's sort of theory-theory of mind and his inclination to conceive of experience as a form of theorizing and experiential states such as emotions, perceptions, and intentions as theoretically postulated entities (Zahavi 2007, p. 23). In short, beliefs, which play a decisive role in Dennetian heterophenomenology as data of scientific inquiry, are not at all primary sources but secondary items of experience. Is there a dimension of experience unbeknownst to the heterophenomenologist? The point would be to answer yes and to show how this dimension is adequately brought forward in phenomenological philosophy. This is the task to be performed in the remainder of this chapter.

According to Dennett's conception of worldly experience as a theorist's fiction (1991, pp. 78-81), phenomenology understood as autophenomenology obeys not to something completely false or utter farcical and it has to be acquiesced as a dimension, albeit unreliable, of everyday human practice. But the private and intimate subjective experiences everybody claims to have that stem from the autophenomenological dimension constituting people's folk psychological views on their own worldly experience, must be submitted to a principle of 'metaphysical minimalism.' This principle "begins by cautiously saying nothing at all about what consciousness might be, or even where it might be found" (1982, p. 159). It is curious to observe that, for someone who claims to have gotten rid of the

'Cartesian Theatre' (Dennet 1991, p. 101 ff.), all these precautionary measures resemble the ones once took by Descartes himself. Dennett echoes then the Cartesian leitmotiv of caution and doubt and the search for scientific clarity and distinction, for one could be deceived even in the most straightforward of perceptions. Indeed, it must be acknowledged that often the apparently commonest experiences are not always what we think they are and that certain aspects of those experiences would have remained unbeknownst to us forever, were it not for the commendable workings of scientific research.

For Dennett, the counterpart of human experience is to be detected in its being expressed in beliefs and opinions of all sorts. This dimension has been traditionally conceived of as the subjective character of consciousness (for instance, in Descartes' use of the first-person singular in his philosophical soliloquies, Dennett [1991], p. 66), which by necessity involves the first-person point of view. This point of view comprises all those experiences accompanied by an authoritative conviction and intimate feeling of what is so experienced and perceived. But this pretension of the agent of having full authority over her own private experiences, is one—as of yet, nonproven albeit manifest—certitude, Dennett warns against. There is, indeed, a difference to be made between what is happening in the agent and what *seems* to be happening: "you are *not* authoritative about what is happening in you, but only about what *seems* to be happening in you, and we are giving you total, dictatorial authority over the account of how it seems to you" (1991, pp. 96-97). For Dennett, this is precisely where one encounters the fundamental mistake of autophenomenology: in its gullible insistence that what seems to be happening in the private, subjective level of experience is to be given entire trust. Soldati has argued, "if experiences are seemings, then… the subject has privileged access to its own experiences" (2007, p. 95). But if experiences are not mere seemings but instead neural events creating those seemings, then the subject has no privileged access to them. The point is that there is more to the seeming dimension of experience. Autophenomenological introspection tends to acquire a sort of infallible stance by means of which it can claim full authority over incorrigible and ineffable experiences. But, of course, no scientific endeavor can be built upon people's folk psychological views on their subjective experience. To repeat Dennett's mantra again: folk psychology *as* theory is mere false ideology.

Dennett takes then a further step and puts this autophenomenological introspection with its purported authoritative incorrigibility at the same level of phenomenological philosophy, by ascribing to phenomenology a methodological search for inner life and private musings:

> Perhaps when people first encounter the different schools of thought on phenomenology, they join the school that sounds right to them, and each school of phenomenological description is basically right about its own members' sorts of inner life, and then just innocently overgeneralizes, making unsupported claims about how it is with everyone (1991, p. 67).

Again, Metzinger agrees almost *verbatim* with Dennett in ridiculing phenomenology for the absurd pretension of generating data by mere invoking first-person judgements. Phenomenology, argues Metzinger, could lead perhaps to arcane disputes when arriving at conflicting statements such as the following:

> "This is the purest blue anyone can perceive!" versus "No, it *isn't*, it has a faint but perceptible trace of green in it!" or, "This conscious experience of jealousy shows me how much I love my husband!" versus "No, this emotional state is not love *at all*, it is a neurotic, bourgeois fear of loss!" (2003, p. 591)

In the same vein, Dennett reduces phenomenological experience and subjectivity to something merely private and ineffable, which, just like for Metzinger, would necessitate a foolish case for settling conflicting views, jeopardizing from the outset the scientific enterprise of arriving at intersubjective and public truths. This is, no doubt, the reason sustaining Dennett's parlance of personal phenomenologies: my phenomenology, your phenomenology, etc., which amounts to everyone's view on what it seems to them to be under a certain mental state of experiential affairs.[81]

As has been asserted several times above, Dennett does not deny the existence of personal phenomenologies, or perhaps better construed from Dennettian jargon, autophenomenologies: "am I saying we have absolutely no privileged access to our own conscious experience? No, but I am saying that we tend to think we are much more immune to error than we are" (1991, p. 68). There is something to be done with all these kinds of personal experiences we all seem to have and which populate our mindful beliefs and opinions. And this is where heterophenomenology comes to our aid, precisely in dealing with the problem of generating genuine scientific data which first and foremost appear to belong to the domain of autophenomenological experience. Thus, let us be clear: neither

[81] The same terminological lack of rigor that was observed in Dennett's case when using the term 'phenomenology' (which sometimes is taken to be folk psychology, sometimes just introspection and on occasion phenomenological philosophy) can be imputed to Metzinger. See Zahavi's objections on this regard (2005, p. 11).

Dennett nor Metzinger, who *prima facie* distrust the status to be bestowed upon first-person experience, are willing to take autophenomenolgy off the list of problems which have to be dealt with when giving an account of consciousness. After all, the phenomenon according to which it seems to us that we are perceiving something, feeling pain, and the like, exists, so there is no point in denying that very fact. What must be done is to acquire "all the heuristic power from first-person descriptions without being driven to naïve realistic assumptions and the stipulation of mysterious, nonpublic objects" (Metzinger 2003, p. 591). As Dennett has argued on his own account, the problem is not that people have these kinds of experiences but the attempt, doomed to failure, to strive for making a theory out of those experiences and ascribing philosophical powers and quasi magical properties to them: "I deny that there are any such properties. But I agree wholeheartedly that there seem to be" (1991, p. 372). This is also the reason why Dennettian heterophenomenology enters the stage in order to salvage phenomenology, construed as autophenomenology, from its Cartesian inherited prejudice. Heterophenomenology is then "the *neutral* path leading from objective physical science and its insistence on the third-person point of view, to a method of phenomenological description that can (in principle) do justice to the most private and ineffable subjective experiences, while never abandoning the methodological scruples of science" (Dennett 1991, p. 72).

It is, after all, a method. In which way does it procede? It bears reminding that the modest goal Dennett had set himself from the start with heterophenomenology was meant as the beginning of a science and not its end. Thus heterophenomenology just "is the organization of the data, a catalogue of what must be explained, not itself an explanation of a theory… And in maintaining this neutrality, it is actually doing justice to the first-person perspective" (2003, p. 27). Dennett's theoretical scruples are meant to redeem all those mental states we commonly associate with the first-person perspective for the purposes of scientific research by way of submitting them to a concrete and objective method. Dennett assumes that by means of paying heed to isolated subjects' reports—isolated inasmuch as they are reporting about certain phenomena that they are confronted with in laboratory-rat-like environments specially designed for that purpose—the scientists dealing with consciousness could in principle arrive at reliable conclusions about the mental phenomena subjects claim to experience, but

this would be just a categorization of beliefs and nothing more. It is a categorization of what subjects merely claim.

So this is how it goes: in order for mental phenomena to overcome their insusceptibility to empirical science, subjects' beliefs about these phenomena must be submitted as reports which, on the other hand, can be categorized by the heterophenomenologist: "we take some of the noises and marks made by subjects as consisting of communication—oral and otherwise—and compose transcripts, which then are further interpreted to yield an inventory of speech acts, which are further interpreted as (apparent) expressions of belief" (2006, p. 37). The heterophenomenologist must maintain her neutrality at all times, that is, she does not judge whether any subject's reports are to be assessed as illusory, imaginary or whatever. On Dennett's view, these heterophenomenological reports are not *about* some outer data, which later on could magically find a consciousness correlate. He is not interested in what the *seemings* of those reports are about, but in the reports themselves, in what the subjects claim, because the reports, in fact, are the data that have to be explicated: "the *primary* data are the utterances, the *raw*, uninterpreted data" (Dennett 2003, p. 21). Thus, the heterophenomenologist goes on to render an account of these reports and utterances which, at least for the time being, are equivalent to pretheoretical data, and subsequently—that is, when finally interpreted heterophenomenologically—they will form the body of categories constituting the heterophenomenological research field: "this transformation of the raw data of acoustic pressure waves, lip-movements, button-pressings and such into expressions of belief *requires* adopting the intentional stance. It requires us to treat the subjects as if they were believers and desirers capable of framing and executing speech acts with intended meanings" (2006, p. 37).

Nonetheless, as we may suppose, a catalogue of beliefs—albeit carefully interpreted—is not a constituent of a well-developed science. This explains perhaps Dennett's insistence in resorting to fiction. It bears reminding that Dennett's treatment of beliefs from the intentional stance as theorists' fictions (1971, 1987 and 1991) is part of the so-called *neutrality* of heterophenomenology. Heterophenomenological reports get to be believed as if they were part of a fictional world. But that does not mean they are plainly false. Actually, "the subjective world is not to be confused with the real world, but that does not mean that it is not by and large composed of truths" (Dennett 2007, p. 262).

As there are most certainly truths in David Foster Wallace's *Infinite Jest* or in Robert Musil's *Der Mann ohne Eigenschaften*, some of the heterophenomenological reports might turn out to be accurate some day, when we learn first—with the help of the heterophenomenological method—how first-person data can be 'reduced' to third-person data (Dennett 2006, p. 149).

Let us now turn briefly to Metzinger's Self-Model Theory of Subjectivity, which develops the author's story about precisely what properties representations in a given information-processing system must possess in order to become phenomenal. Metzinger's position is clearly counterintuitive for it argues for the elimination of the notion of self. Human agents believe they are someone, and they experience themselves as such, but the feeling of being someone is literally phenomenal in that it is only an appearance. In chapter 6 of *Being No One* (2003, pp. 299 ff.), Metzinger presents his idea that human agents are constantly confusing themselves with the content of their phenomenal self-model. Conscious systems are presentational since they are endowed with the capacity of generating an internal global depiction of parts of reality. What is most striking is the fact that human agents can also generate a representation of themselves. Part of the system is a sort of 'inbuilt blind spot'—Metzinger calls this phenomenon 'autoepistemic closure' (2003, p. 57)—which explains why human beings are not able to realize that the content of their subjective experience is actually representational. In short, complex biological organisms like human beings possess a conscious self-model but they are not selves, just highly complex brain states (2003, p. 563). Human agents are caught in a naïve-realistic self-misunderstanding, from whence all their rich phenomenology can be deduced (2003, p. 332). Astonishingly, for Metzinger, being conscious during the waking state can be characterized as an online hallucination (2003, p. 51).

What these approaches construing phenomenal consciousness as illusion (Dennett) and confusion and overall hallucination (Metzinger) might concede to phenomenological philosophy is perhaps a modest contribution, which could eventually enrich "the vocabulary of the personal level…, teasing out aspects of the patterns of competences, inabilities, needs and methods of persons in illuminating ways, but this is all just setting the specs—the competence model—for the subpersonal level account of how the performances are achieved" (Dennett 2007, pp. 256-257). However, phenomenological philosophy is not in the business of enriching the vocabulary of our subjective

beliefs. The next section of this chapter, exhibits how phenomenology tackles the problem of consciousness.

Consciousness and the Natural Attitude

The discussions presented until this point can be seen as proof that the term 'phenomenology' is ubiquitous in consciousness research, the philosophy of mind, and the philosophy of psychology. However, the potential of phenomenological philosophy and its transcendental (antinaturalistic) stance has barely been exploited. And this, due to the terminological confusion by means of which phenomenology has been equated with an internal perspective (*introspicere*) dealing with occult objects like qualia or with an idiosyncratic point of view about a series of mental events. What do we get from contrasting real existing phenomenology (Soldati 2007) with autophenomenology? Not only, as we would like to show, a quite distinguishable philosophical enterprise from the extravagant overgeneralization out of intimate, ineffable musings Dennett ascribes to phenomenological philosophy, but above all a whole dimension of experience which is altogether left unattended on behalf of heterophenomenology's neutrality and recourse to beliefs as data.

But first, let us get things straight from the phenomenological point of view concerning introspection. Is it accurate to vindicate phenomenology as a method "based on a special technique of introspection" (Dennett 1991, p. 44)? Is it true that phenomenology has been committed to a form of "methodological solipsism" (Dennett 1987, p. 154)? In a nutshell, is phenomenology plain autophenomenology? For anyone even slightly familiar with the phenomenological tradition, the answer to these questions has to be an emphatic *no*, if by phenomenology one understands the classical phenomenological movement founded by Husserl and its ensuing contemporary heritage. As Zahavi and others have argued in several places (2007, 2008 and Gallagher & Zahavi 2010), the claims according to which phenomenological philosophy admits of a facile equation with such "introspectionist bit of mental gymnastics" (Dennett 1987, p. 153), only end up revealing "one's lack of familiarity with the tradition in question" (Zahavi 2007, p. 28; Gallagher & Zahavi 2010, p. 21).

Husserl himself, to be sure, has rejected in more than one occasion the anodyne confusion between inner perception (*innere Wahrnehmung*) and phenomenological perception (*phänomenologische*

Wahrnehmung) (Hua XXIV, 216). In *Ideen III*, precisely when discussing a contemporary misrepresentation of phenomenology that Husserl conceived of as superficial (*oberflächlich*) and preposterous (*grundverkehrt*), he attacks the view that phenomenology could be assessed as the "*Restitution der Methode innerer Beobachtung*", that is, as the restitution of the method of introspection, which would be supposedly in charge of "*direkte innere Erfahrung*" or direct inner experience (Hua V, p. 38). And not only that: the very point of departure of phenomenology in Husserl's breakthrough work, *Logische Untersuchungen* (1900-1901), was precisely a call to abandon the dichotomy (*Scheidung*) between inner and outer perceptions, which Husserl associated with a naïve commonsensical metaphysics left behind with the concept of intentionality (Hua XIX/2, p. 673). But again, as Zahavi reminds us, this facile divide between inside and outside "is precisely something that the term 'introspection' buys into and accepts" because "to speak of introspection is to (tacitly) endorse the idea that consciousness is inside the head and the world outside" (2007, p. 29). So, at least from the point of view of phenomenological philosophy, introspection happens to be from the very outset antiphenomenological. In the same sense, Heidegger has severe words against a conception of an encapsulated (*verkapselt*) Dasein striving to abandon a purported inner sphere in order to reach the outside world (SZ § 13, p. 62). For both founding figures of phenomenology, such a view clumsily dividing reality has its roots in the 'natural attitude' (*natürliche Haltung* in Husserl) or in the primacy of the theoretical (*Vorherrschaft des Theoretischen* in the young Heidegger); both of which are of great interest to phenomenology, inasmuch as phenomenology also investigates the origin of theory and abstraction and its roots in the life-world.

But shouldn't we judge phenomenology as a rather odd undertaking, specially from the point of view of method? Dennett claims that the phenomenological *epoché* is a special technique "in which the outer world and all its implications and presuppositions were supposed to be 'bracketed' in a particular act of mind" (1991, p. 44). Isn't that just an obscure procedure that annuls reality, objectivity and, even worse, the very possibility of science? This explains Dennett's assessment of Husserlians as self-styled obscurantists involved deeply into abstruseness for its own sake (1994). But if by *epoché* one understands an exclusion of reality, a sort of repudiation of reality or an irrational cancelation of objectivity, then we are dealing here with a quite flawed and nonphenomenological definition of *epoché*. The bracketing of the phenomenological *epoché* is not the attempt to annul reality, but rather to let

without effect the natural attitude pervading even science. In fact, the reason why introspection is a concept with no interest for phenomenological philosophy can be spotted in the mistake of characterizing phenomenological philosophy as a method striving for clarifying reality from the first-person point of view. On the contrary, phenomenologically speaking, both perspectives (the first-person and the third-person viewpoints) conceal their hidden resemblance: they are both objectifying stances having as origin the natural attitude. Husserl's *epoché* (*pace* Dennett) does not annul objectivity but explains it and the bracketing implies cancellation of both the first-person and the third-person viewpoints. The two perspectives overlook the most important aspect of consciousness: that it cannot be turned into an object for it is because of consciousness that objectuality (*Gegenständlichkeit*) itself is possible.

The transcendental stance often, and correctly, attributed to phenomenology means that it understands itself not only as contributing to the growth of scientific knowledge and its positive results, because first and foremost it is interested, as a *philosophical* endeavor, in the ground floor of human knowledge and its conditions of possibility. So, from a phenomenological standpoint, even science and its objectivity cannot be taken for granted; even science gets to be phenomenologically questioned. Above all, the most basic presupposition with which science begins its investigative journey, that it can investigate things objectively, gets to be questioned from the phenomenological standpoint. The world exists without the shadow of the slightest doubt and it is also possible to study objects and found suitable theories that explain their causal relations. In conjunction with the phenomenological *epoché*, the phenomenologist will not deny neither the existence of the world nor the possibility of studying it objectively. But what is interesting and really worth reflecting upon, as Husserl thought rigorously, is to comprehend this certitude and clarify its legitimacy (Hua V, pp. 152-153).

Therefore, against Dennett, here is a phenomenological definition of *epoché*: it is a technical term that refers to the suspension of the natural, realistic, and noncritical inclination to take the world for granted. The subsequent aim of this so-called *Methode der Einklammerung* or 'method of parenthesizing' (Hua III, § 32 p. 56), is admittedly not to neglect or exclude reality but to neutralize a certain dogmatic attitude towards reality, thereby allowing for a focus on the meaningfulness of the world as given in experience. This is, indeed, the discovery of a new scientific domain (*idem*) and not

the exclusion of the entire world at the behest of a solipsistic world appearing just to me. Husserl is very clear about it: "I am *not negating* this 'world' as though I were a sophist; I am *not doubting its factual being* as though I were a skeptic" (Hua III, § 32 p. 56). What is excluded as a result of *epoché* is most certainly a dogmatic naïvety. But this naïvety is even present in science, whose theoretical mission is taking things as objects. But again, objectivity is everything but natural, for it is rather a by-product of certain historical experiences, above all, of intentionality. As Hans Blumenberg has handsomely defined it, Husserlian intentionality is the special mark of consciousness; and the latter meanwhile can be understood, not as something merely mental, encapsulated in people's brains, but as a historical and productive structure (*Leistungstruktur*), striving for completion and comprehensive goals (*Zielstrebigkeit*) (Blumenberg 2009, p. 18). Let us look at this in depth: "the intentional character of consciousness fulfills itself ultimately in the most all-encompassing horizon of horizons, in the 'world' as the most regulatory idea of possible experience… 'Nature' as well is the result of a certain conceptual sharpness and this is why it is to be held as a by-product of, and not as originary as, the world" (Blumenberg 2009, p. 19).

Notice that the natural attitude with its typical objectualist conception of subjectivity is so pervasive that it has rendered the alternative phenomenological explanation of consciousness almost invisible. Rowlands explain this as follows:

> subjective, conscious, phenomena are not parts of a region of reality to which our access is idiosyncratic, and where this idiosyncracy constitutes their subjectivity. Indeed, they are not parts of a region of reality to which we have access at all. Rather, subjective, conscious phenomena are ones that belong only to the *access* itself. There is no region of reality to which conscious phenomena belong, or in which they find their place. Rather, conscious phenomena simply belong to our *accessing* of regions of reality. (2010, p. 89)

From the transcendental standpoint of phenomenology follows correctly that reality is not intrinsically objective (which means only that reality is more than objectivity, not that reality admits of no objective description). In order to explain the reason of why this is the case, phenomenology remarks the hybrid character of consciousness: its capability of being both *act* and *object* of experience. Certainly both act and object can be objectified, for I can look at the thing being perceived or at the act of perceiving the object. What phenomenology adds to the traditional view of objectuality is that there is an intentional core in experience that cannot be eliminated or reduced: there is always an aspect that is irreducibly transcendental, that is, not something of which we are aware but something by virtue of which we are

aware. This is the subversive idea of phenomenology: that objectuality must be brought to a halt because the mode of presentation of the object cannot itself be turned into object. Again, "the mode of presentation is not something *of* which we are aware (as we might, for example, be aware of aspects) but something in *virtue of which* we are aware of the intentional object of our experience" (Rowlands 2010, p. 92).

So Dennett wanted to know what in the world could not be investigated by heterophenomenology. Here is the phenomenological answer: heterophenomenology, as a product of the natural attitude, cannot investigate the transcendental mode of presentation which confers meaning to our experiences. What is worth keeping in mind is that phenomenology, whatever its themes and interests, is not looking for a turn to inwardness or for a mental introspective encapsulation. On the contrary, it is the world and the meaningful structures that constitute this horizon of all horizons what phenomenology attempts to investigate. Phenomenology remains an investigation of reality and not of private, ineffable seemings. But because reality is more than the capability of objectifying it inasmuch as reality is meaningful due to the access to it provided by consciousness, it is nothing but a category mistake (see Lembeck 2010, p. 175-176) to conflate consciousness with another object of physical reality. So this fundamental insight of phenomenology can be directed critically to the positions on consciousness (Chalmers's, Searle's, Dennett's and Metzinger's) that have been already dealt with.

Dennett's extreme caution has not prevented him from precisely, and contrary to his own intentions, not doing justice to subjective experience and overlooking central aspects of it. And this is what one gets when one assumes that beliefs could count as originary and not—as it should be—as merely derived from experience. But, of course, as phenomenologists, we should not expect this to be a dogma, as it is the *belief* that beliefs should occupy the place Dennett ascribes to them. On the contrary, it should be phenomenologically evident and not sheer stipulation by fiat. Indeed, phenomenology means basically this: *Nothing can be made up!*[82] On the contrary, for Dennett, "subjective, then, are those experiences that are only the object of a belief" (Soldati 2007, 96) and these are the only experiences that Dennett, as heterophenomenologist, is willing to accept. But is this the case? Do beliefs exhaust all possible kinds of experiences? Dennett overlooks precisely (and conveniently for his theory) the

[82] This is a way of interpreting Husserl's *Prinzip aller Prinzipien* (see Hua III § 24).

nonpsychological dimension of consciousness: the very dimension of givenness and its structures. Objects appear to us in our very human world, which means that we are as subjects also datives of the manifestation of reality. So the sense of reality will not be disclosed if one supposes that all there is to reality is unworldly abstract objects. As Ratcliffe argues: "in order to reflect upon the nature of our sense of reality, a very different kind of enquiry is required, one that seeks to make explicit those aspects of experience that are ordinarily taken for granted and to study their structure. This, amongst other things, is what phenomenology aims to do" (2007, p. 492).

Belief is a no starter in philosophy because it objectifies from the outset and takes things as mere thinghood. Reporting on our experiences is a matter of not being absorbed by a certain situation but just a belated reflection upon it. Careful description of the meaningful dimensions of the life-world (Husserl) or of existential understanding (Heidegger) constitutes the task of phenomenological investigation. But this meaningfulness given in experience or this existential understanding, which gives form to our genuine way of 'knowing,' is not mental at all, let alone inward introspection: "this 'knowing' *does not first arise from an immanent self-perception* but belongs to the Being of the 'there,' which is essentially understanding. And only *because* Dasein, in understanding, is its there, *can* it go astray and fail to recognize itself" (Heidegger SZ § 31, 144. Emphasis added). Phenomenology's task of describing these meaningful structures of experience is essential, also for understanding the origins of objectivity.

In addition, there seems to be an inadmissible rejection of perception in Dennett's heterophenomenological method. On the contrary, Husserl demanded paying a great deal of heed to perceptive experience in his *Prinzip aller Prinzipien* (cf. Hua III § 24, 52), which therefore is to be given a paradigmatic stance. To begin with abstraction in order to explicate meaning is a no starter for phenomenology, because even objectivity derives its legitimacy from an intentional understanding which precisely allows the taking of things as objects. In the same vein, supposing beliefs are to be bestowed with an originary character begs the question, for here one assumes without proof what is to be proven. Imposing belief by sheer fiat *à la* Dennett, that is, not phenomenologically, as the first stance in the investigation of subjectivity, presupposes actually what it has to be proven: whether verbal thematization and circumspective absorbed coping are to be conceived on the basis of the former's primacy over the latter or conversely. Beliefs are only possible on the basis of a first perceptual

encounter with states of affairs. The world in which objects appear is meaningful from the start and by no means can be reduced to a collection of abstract entities being postulated by linguistic declaration. In this very sense, phenomenology does not annul objectivity but construes it as a *terminus ad quem*, not as *terminus a quo*. As resulting by-products of pretheoretical experience, objectivity and abstraction do not stand on their own feet and can never render an account of themselves, for they belong to the meritorious, albeit constantly overlooked, workings of circumspective perception. And the same counts for thematic beliefs. They appear when obviousness in the life-world is interrupted by malfunction, which demands immediately a theoretical and thematic treatment of the situation. Indeed, it is possible to make a theme almost about anything. But in order to be able to make a theme about anything there is always a mode of presentation which cannot be made thematic, since it expressly makes possible that something at all appears as thematic.

So is consciousness an illusion? One tremendous fact of the pervasiveness of the transcendental standpoint is that even an illusion presupposes a mode of presentation. Therefore not even the illusory or hallucinatory character of consciousness suggested respectively by Dennett and Metzinger could be investigated without a vehicle that makes possible such presentation of reality. This means that the fundamental character of consciousness is *disclosure*. And such disclosure is only possible over the basis of an unified field of phenomena, which Heidegger calls *Entdecktheit* (GA 20, p. 349): the character of *discoveredness* of the world which allows in the first place for such disclosing. Any given experience contains this meaning-giving aspect of intentionality: that it discloses a world for us. Trying to explain this disclosure by the objects that it allows to appear, that is, trying to explain empirically what in itself is nothing but the fundamental presupposition for every empirical theory (see Windelband 1909) is as absurd as trying to justify the scientific claims of Einstein's theory of relativity by dissecting and analyzing his brain (see Lembeck 2010). This is exactly what Husserl meant (see Hua XXV, p. 9) when he warned about the absurdity of a theory of knowledge based on natural science as well as any psychological theory of knowledge implying there is nothing more to consciousness than the explanation of natural events.

The hard problem of consciousness gets its 'hardness' at the behest of an impossibility which is not recognized as such: namely, that consciousness itself cannot be naturalized, nor reduced, nor

objectified. It is the facile equation of consciousness with just another object that appears in the natural world what causes all the trouble. Phenomenology, on the contrary, judges nothing regarding natural lawlike occurrences but only the modi of consciousness characterizing meaningful experience. As Husserl has very radically put it, not even if nature were a phantasm or objective science sheer craziness would phenomenology, nor its justifiable entitlement for research, suffer the slightest bit from it (Hua XXV, p. 90).

The Really Astonishing Hypothesis

What moral should be drawn from the reflections comprising the guiding thread of this discussion on consciousness? The standard reaction to the claim that phenomenological philosophy might contribute to the discussions customarily framed in the philosophy of psychology and the philosophy of mind is the complacent acknowledgement that it might have 'good ideas' regarding the nice elements that populate the wonderful 'phenomenological garden' (see Dennett 1991, p. 374). This is certainly a concession to eliminativism, since Dennett's strategy is to attempt to break the spell of the phenomenological garden. The garden, along with the experiences and feelings that populate it, is just the seeming of a magic trick put into operation by a complex (and anonymous) series of mechanisms, which do not see, do not feel, and do not perceive. The billions of minuscule robots doing the job might be alive, but they certainly do not exist, since existence is also a 'feel' which, for Dennett, admits of elimination.

Other reactions to the phenomenological challenge are not less scornful. Take for example Metzinger's (2005) reply when he was confronted by Dan Zahavi (2005) about a series of unjustified philosophical assumptions in his book *Being No One*. Metzinger graciously thanked being informed about a 'beautiful sentence' by Michel Henry and implied Zahavi employed 'beautiful phenomenological poetry' (2005, p. 3). In this context, Metzinger posed the following question: "where is the positive, *systematic* contribution of German phenomenology to the issues the global philosophical community faces today? Where is the *phenomenological* contribution that lives up to the standards of conceptual precision of today's best current philosophy of mind?" (*idem*). However, the concept of consciousness has been everything but precise. In addition, not only the conflation between the

empirical and the transcendental has been overtly ignored by Metzinger in his book, but neither the mereological fallacy in neuroscience (that brains have experiences, make guesses, interpret clues on the basis of information, categorize, construct hypotheses, estimate probabilities and the like, see Bennett & Hack [2003], pp. 68 ff.) has been touched on by him even indirectly. Does this forgetful treatment of the assumptions underlying Metzinger's work meet the standards of conceptual precision? As for the 'positive' contributions of phenomenology, there are none to be made, for philosophy is not empirical science and cannot render positive results unless it wants to relinquish the philosophical standpoint. It would just be due to an 'identity confusion' or an inferiority complex that phenomenology would like to abandon its most proper stance.

But what is this stance? The answer can be illustrated by referring a fact that has been considered astonishing: precisely, that all that we are as human subjects is encapsulated as information in that box of gray matter inside our skulls, the brain. In *The Astonishing Hypothesis* (1994), Francis Crick affirms that "what you see is not what is really there; it is what your brain believes is there" (p. 30). But as Noë (2002) has argued, the complete opposite is what would be astonishing: that consciousness does not occur inside the skin of human beings, in the brain. And this because internalism was actually the Cartesian idea that humans are identical to an interior something whose essence was consciousness (p. 5). So the contribution of philosophy to the more vexatious and difficult questions of the present is not positive, but critical. Philosophy not only contributes to conceptual clarity but also to the disentanglement of traditional claims that are taken for granted without being made explicit and whose conceptual history are wholly ignored. More crucially, philosophy helps in clarifying that science is also premised on the basis of assumptions, many of which are of philosophical heritage. Philosophy can show that science did not simply descended from heaven, but that it is also a product of human history. This might sound as anathema for philosophers clinging to scientistic views, but science is human, all too human. So from philosophy too one can learn which questions need to be posed and which not.

As regards the thematic field which has been explored in this chapter, phenomenology renders a nonnegligible contribution in making clear that naturalizing consciousness constitutes an in-principle impossibility. This contribution is critical because it establishes limits to what can or cannot be investigated from the objective perspective. This, however, invites the obvious retort that science is

autonomous and is not in need of listening to philosophical sermons about what can or cannot in-principle become part of a research program. But critical talk as regards limits in this case is also a way of marking out the proper place of philosophical inquiry in the heyday of scientific hegemony. Philosophy has quite a different task to pursue than the one sought in the natural sciences.

From the phenomenological standpoint, consciousness is precisely not amenable to the objective perspective that characterizes empirical research. The character of consciousness is to be understood as fundamentally coupled with meaningful acts, and not with mere logical rules or abstract contents. The intentional possesses a vertical structure in that the directedness of consciousness is also a lived experience of a mode of presentation. Mental contents, to use the analytical vocabulary, are meaningful because they happen to someone, they make their appearance to someone, and are perceived by someone. A *noema* is a meaningful phenomenon and not just a by-product of computational operations that could be indifferently interpreted by a machine lacking existential situatedness. This amounts to saying that a presentation of meaning cannot be captured by way of a series of algorithmic rules. Not even if billions of commendable mechanisms were aptly explained, would it be possible to stumble upon semantics. Meaning is only captured in an act (*noesis*) that is both lived and experienced. This is what phenomenology has brought to the fore: that this dimension of meaning is not a matter of being represented, but instead of being lived with understanding by someone. So what is even more embarrassing for a computational and mechanistic approach is not that it has reduced consciousness to something more fundamental, but that it has overlooked consciousness in its entirety while at the same time *supposing* it.

On a final note, let us reflect upon the really astonishing hypothesis, which is not that human beings are highly complex machines—an idea so traditional that it has been on offer for centuries—or self-models that cannot recognize themselves as such models. As Metzinger affirms about the model: "it is transparent: you look right through it. You don't see it. But you see *with* it. In other, more metaphorical, words, the central claim of this book is that as you read these lines you constantly *confuse* yourself with the content of the self-model currently activated by your brain" (2003, p. 1). In contrast, this is how the issue can be phenomenologically stated: what is transparent is, in fact, consciousness. You see with 'it' but you cannot make a theme out of 'it' because 'it' is no 'it.' Consciousness is not an

epistemological machine with special powers for generating a virtual qualitative illusion. Consciousness is rather factic-life experience itself, which must be differentiated from a merely *kenntnisnehmende Erfahrung* or theoretical take on things. As Heidegger points out: what is so lived [*das Erlebte*] is world, that is, no object at all (GA 60, p. 11). Incidentally, Wright was on the right track when he claimed that "*objectivity* can never be equated with *existence*" (2011, p. 352). But if it is possible to make a theme out of consciousness, thus transgressing the fundamental impossibility of treating consciousness as an object, that is only because of the pervasiveness of the natural attitude. The illusion is not that meaning be taken as real, but rather to take as object what makes possible the presentation of all meaning, including objects.

| 8 |

HEIDEGGER AND COGNITIVE SCIENCE

The Heideggerian Alternative

Upon reviewing a half a century of AI as a research program and some of its theoretical challenges, Margaret Boden (1995) notes that neo-Heideggerian murmurings are now afoot that threaten "the fundamental assumptions of AI, for they reject the subject-object distinction presupposed by realists and idealists alike, and deny the epistemological primacy of science" (p. 99).[83] As though AI were not rich enough in perspectives and disciplines (what certainly has contributed to its being plagued with disputes and fundamental debates of cross-disciplinary nature), the overall picture is complicated by the fact that now phenomenological approaches are to be taken into the equation. What is more interesting is that the confirmation of this ongoing reception of phenomenology in cognitive science should let us devise the contours of an alternative philosophy of science, which is to be forthwith explained.

According to Boden, the chief difference of these new approaches having phenomenological (mostly Heideggerian) leanings with Dreyfus's, for example, is that they are perceived as bringing to the fore not only criticism—like Dreyfus's incorporation of phenomenology in cognitive scientific discussions—but practical solutions for a series of concrete problems in technological implementation. The idea here is that a series of new assumptions can contribute to carry out research that can illuminate the nature of cognition, as though the founding purpose of cognitive science could be realized if it is now premised on a different theoretical framework. On this view, Dreyfus was just a staunch critic. For their part, the researchers partly or wholly inspired by the critic must bring forth proposals in lieu of mere negativity—as was Dreyfus's primary intention. As Boden (1995) has argued, the Heideggerian approach is compelling since it is now

> being mounted by people sympathetic to computer modeling: in particular, to situated and evolutionary robotics and to A-Life's studies of 'animats.' These people see organisms as dynamic systems closely coupled with their environment. Instead of positing internal

[83] 'Neo-Heideggerian' refers particularly to Michael Wheeler's approach (see 1995), who has proposed some principled modifications to Heidegger's own ideas so that they can be put to work in cognitive science. However, this specific approach shall be dealt with later in chapter 9.

> representations of an objective external world, they speak of whole systems embedded in, and adapted to, their own particular 'worlds.' (*idem*)

In this regard, the reception of phenomenology within cognitive science is very significant and, for that matter also, very surprising. First, one needs only reminding that a great majority of researchers working on the basis of the analytical tradition of philosophy have always been keen to assess European or 'continental' philosophy (whose founding movement is phenomenology), at its worst, as obscure, muddled, antiscientific, and even a little zany. At its best, phenomenology is construed as idealism and historicism: always preoccupied with the clarification of some immanent jargon and the sheer exegesis of texts.[84] And second, it is certainly more perplexing to note that it is Heideggerian philosophy the one which has been brought to bear on how cognitive science might help unblock some of its more recalcitrant stalemates. After all, it was Heidegger who said that science is incapable of thinking (in his famous formulation that *die Wissenschaft denkt nicht*), for it does not reside in the same dimension of philosophy, although it entirely relies on it (GA 16, p. 705; GA 8, p. 9). As Heidegger was quick to point out, "*no science can know from itself about its own fulfilled form of knowing*" (BW, p. 12). This is Heidegger's idea of an immanent limitation belonging to science: that its own method withholds the reflection upon itself.[85] So with this in mind, let us revise some of Heidegger's ideas on science, about which one cannot simply remain noncommittal when trying to understand to what extent a Heideggerian cognitive science even makes sense.

According to Heidegger, the idea that science does not reflect upon itself is not to be understood as a deficiency which could somehow be alleviated by means of incorporating more

---

[84] See, for instance, on Heidegger's 'jargon,' Adorno (1964) and Bourdieu (1988). Oxford philosopher of information, Luciano Floridi, has recently claimed there is a great sector of current philosophical research haunted by a species of scholasticism which, "as an intellectual typology rather than a scholarly category, represents a conceptual system's inborn inertia, when not its rampant resistance to innovation" (2012, p. 9). On Floridi's account, this is institutionalized philosophy at its worst, which can be deemed "a degeneration of what sociolinguists call, more broadly, an 'internal' discourse of a community or group of philosophers" (*idem*). According to such claims, it is not difficult to imaging phenomenology as belonging to such philosophical communities that adhere to some discourse set by a particular group of philosophers and founding texts. On a similar line, Schnädelbach (1981) has criticized a sort of 'being-towards-the-text' that characterizes great part of contemporary philosophical practice. This is the *morbus hermeneuticus* (for Schnädelbach, a philosophical sickness): the idea that "philosophizing consists in reading the work of philosophers and that philosophy takes place where philosophical texts are interpreted" (1981, p. 3). For phenomenologists, however, the task at hand is not to clarify texts but reality.

[85] "Hier zeigt sich die *innere* Grenze der Wissenschaft: ihre eigene Methode versagt in der Besinnung auf sich selbst" (BW, p. 12). Notice, however, that Heidegger is here elaborating on an ample definition of 'science,' which also includes the *Geisteswissenschaften*, or the humanities and social sciences. This explains why, say, not only physics cannot reflect upon itself by means of the physical method or, for that matter, mathematics or geology, but also, for instance, philology cannot reflect upon itself philologically (*idem*).

conscientiously in scientific endeavors the obligation to do that sort of reflective work, nor by founding the usual institutes and programs devoted to the history and philosophy of science. The nature of the reflection to be found, for instance, in the philosophy of science—as is currently known: as the field devoted to analyzing the character of scientific explanation, its theories, and justification claims—is metascientific inasmuch as it is concerned with the examination of the nature of scientific problems and procedures and, moreover, with an analysis of that very examination, and not simply with the mere description of the production of scientific results. In its more extreme objectifying form, metascience can be defined as a theoretical rationale "for studying scientific episodes in order to assist in the understanding and integration of the massive historical track record" (Faust & Meehl 2002, p. 185). The history of science is thus viewed as a sampling of scientific episodes or the historical track record piled up in the historical data base, in order to address long-standing questions in the philosophy and history of science. As such it should assist scientists in higher level and complex integrative judgements (Faust & Meehl 2002).

However, it most certainly goes without saying that such metareflection will not do and will not count as reflection, on Heidegger's terms (that is, as *Besinnung*), because if science reflected upon itself in the way Heidegger demands, it would stop being science (at least, as it is now conceived). As a matter of fact, the usual way of practicing the philosophy of science is what Heidegger would call an *ontic* treatment of science, concerned with its factual development and goals (SZ § 69, p. 356-357).[86] But it is the condition of modern science—in its *Technisierung* and *Spezialisierung* (BW, p. 8) and in its being undertaken by the industry—not to reflect *ontologically* upon itself. This is indeed how science progresses and its inner threat (*innere Bedrohung*) consists in that it flourishes nowadays with a success never seen before (BW, p. 7). So Heidegger does not belie the success of science but rather has it that this precisely is evidence that science does not stand in the truth anymore, having therefore forfeited the grounds of *das wesentliche Wissen* or essential knowing (*idem*). This threat manifests itself in that in science a tranquility is sought, which hence disregards the unrest of questioning (BW, p. 11). Now, this essential knowing which Heidegger imputes to philosophy is no romantic metaphysization presenting us with

[86] Typical questions (here taken from Bechtel [1988, p. 1 ff]) of the philosophy of science understood in this ontic sense are: what is scientific explanation? To what extent can scientific claims ever be justified or shown to be false? How do scientific theories change over time? What relations hold between old and new theories? What relations hold, or should hold, between theoretical claims developed in different fields of scientific investigation?

occult entities unknown to science.[87] It presents us with no entities whatsoever. The point is, strictly speaking, that the sciences investigate entities and not the being of entities, which means that "it is thus essential to a science precisely not to think, that is, not to question being, but rather to proceed with the investigation of its objects" (Glazebrook 2000, p. 217).

So this immanent constraint of nonreflection pervading science must be taken as a confirmation that there lays dormant, as it were, another possible way of knowing that does not yet unfold, which is also pertinent to science and which must be revivified, "assuming that science should have a self-awareness that corresponds to its own essence" (BW, p. 12). As Guzzoni (2012) has argued, Heidegger's critical treatment of science is always carried out at the light of contrasting it (for example, the description of science as being one of the cultural products of humans) with a different viewpoint: "techno-scientific thinking is not and should not be the only kind of thinking… the crucial thing is to contrast it against a different thinking that he calls *besinnlich*" (p. 193).[88] Without the aforementioned self-awareness that Heidegger demands concerning what science is, science cannot know what it wants and ignores thoroughly the direction it should be taking with regard to its tasks. Science can thus make itself indispensable by its practical effects and thanks to its undeniable success, but it can never formatively and educationally contribute, as a spiritual and cultural reality (*geistige Wirklichkeit*), to world history (*idem*). Roughly speaking, science needs philosophy in order to reflect upon itself, since *that* kind of reflection upon science is not scientifically possible (BW, p. 13). Heidegger's point here is, as Rouse argues, that science needs philosophy in order to remain 'in the truth' (2010, p. 180).

Heidegger's conception of modern science as dangerous and threatening is grounded on what he sees as the contemporary complex conjoining technoscience with industry and politics (see Pöggeler 2000, p. 14): precisely the sort of complex behind AI research joining together economic means to do the research coming from corporative institutions, and military implementation under political direction. This institutional character of science contributes, on the one hand, to the objectification of

---

[87] Remember that Heidegger is not extolling the virtues of philosophy over against the defective character of the sciences. "*Das Bedenklichste ist*" says Heidegger, "*daß wir noch nicht denken*" (GA 8, p. 9). As Glazebrook points out, "Heidegger's view is not that the sciences are somehow deficient in comparison to philosophy, but rather that neither science *nor philosophy* is thinking in the modern epoch" (2000, p. 215).

[88] See Heidegger's words in 'Wissenschaft und Besinnung' (1953): "Solange wir die Wissenschaft jedoch nur in diesem kulturellen Sinne nehmen, ermessen wir weder die Herkunft noch die aus dieser verfügten Tragweite ihres Wesens" (GA 7, pp. 37-66).

research, and to the disappearance of the old scholar who is now replaced by the technologist (*Techniker*). On the other, "the researcher needs no longer a library at home. He is, moreover, constantly on the move. He negotiates at conferences and collects information at congresses. He commits himself to publishers' commissions. It is publishers who now determine which books need to be written" (GA 5, p. 85). The scholar is no longer paragon in the culture dominated by technoscience; the expert is. The expert behaves in such a way that she is almost indistinguishable from a businessman contributing to both scientific development and economic growth; tasks which are also identical.

Leaving the political aspects aside (although they belong fundamentally to the equation), one would be advised to pay heed to the consequences of the transformation of science into research in this Heideggerian narrative, given that technoscience allows for entities to appear as calculable and orderable; what undoubtedly resonates pretty well with the demands of the industry. As Heidegger pointed out on this regard: "nature reports itself in some way or other that is identifiable through calculation and that it remains orderable as a system of information" (GA 7, p. 23). Such entities, with which the technologist presents us, are shorn of their practical involvement and thus merely *vorhanden*. But it must be noted that this presentation of things as being 'merely there,' stripped of their original contextual understanding, permits one to inquire about the nature of science itself in the so-called Age of the World Picture.

In *Sein und Zeit*—which incidentally is the most referenced Heideggerian work (specially its first division) by AI researchers—Heidegger showed how the theoretical discovery of present and occurrent (*vorhanden*) entities comes about by modifying our practical involvement in the world. On Heidegger's account, an investigation of the theoretical attitude pertains to the problem of giving an existential account of science, which conceives of science as a mode of existence and as a way of being-in-the-world. In other words, this is the question of the ontological genesis of the theoretical attitude: "we are asking which of those conditions implied in Dasein's state of being are existentially necessary for the possibility of Dasein's existing in the way of scientific research" (SZ § 69 b, p. 357).

For Heidegger, the thematization of objects that leads to objectification in science presupposes being-in-the-world as the basic state of Dasein (SZ § 69b, p. 363). Scientific objectification does not clarify the hermeneutical situation of being. On the contrary, such objectification is only possible on

the basis of a pre-understanding of being which is always assumed: *a world must have been disclosed* to Dasein for it to have any dealings with a specific practical context, even theoretically (SZ § 69b, p. 364). Heidegger's argument concerning the origin of the theoretical attitude shows that there is an important difference to be made between circumspective deliberation (*umsichtige Überlegung*), which illuminates Dasein's factic dealings in particular contexts, and a mere confirmation (*Konstatierung*) of some occurrent (*vorhanden*) objects and their properties (SZ § 69b, p. 359). Circumspective hermeneutic deliberation which brings wordly contexts closer and makes them have a sense, has the existential significance of a presentation (*Gegenwärtigung*), on top of which any representation (*Vergegenwärtigung*) is even thinkable. That is to say, *Vergegenwärtigung* is only possible on the basis of a previous *Gegenwärtigung* enabled by the Open (*das Offene*) which for itself opens up the world (*das Sichöffnende und Offene* [see von Herrmann 1994, p. 213]).

There is also a temporal aspect of any of Dasein's encounters with situations, which has to be taken into account. Situations are *lived through* and, for that very reason, they are not merely theoretically grasped (*konstatiert*). This temporal aspect shows how the theoretical gaze at objects is a *new way of seeing* —and not just 'the' way of seeing *par excellence*—that encounters sheer things present-at-hand (*vorhanden*). Notice that, inasmuch as a new way of seeing which stems from an alteration of the situation, the theoretical attitude can hardly be said to be original. A propositional and objectifying take on things (by means of a sentential assertion like "the hammer is heavy") "is no longer spoken within the horizon of awaiting and retaining an equipmental totality [*Zeugganzen*] and its involvement-relationships [*Bewandtnisbezüge*]" (SZ § 69b, p. 361). That is to say: the grasping of thinghood as the first step allowing for objectivity only takes place when the things that so appear have also been deprived of their temporal dimension. And this is the condition for encountering, for instance, a mere corporeal thing subject to the law of gravity. That is to say, a precondition of abstraction is presenting things as detached and as indifferent to human context. Under this abstract precondition, a thing is *anything*. Its singular place, its particular role in human concerns is annulled. In view of this, says Heidegger, "*the understanding of being* by which our concernful dealings with entities within the world [*innerweltlichen Seienden*] have been guided *has changed over* [*hat umgeschlagen*]" (*idem*). Mere sentential (in contrast with circumspective) talking of something being, for example, 'too heavy' or 'too light,' has no longer any

meaning: "the entity in itself, as we now encounter it, gives us nothing with relation to which it could be 'found' too heavy of too light" (SZ 69b, p. 361). As a philosopher of science, then, Heidegger problematizes the theoretical, objectifying and epistemological stance of science "by tracing their eidetic genesis from their initially pretheoretical and nonobjectifiable matters" (Kisiel 2012, p. 241). What is certainly more remarkable is that these ideas were endorsed by some researchers working in technological research communities. Perhaps they also—so was the mantra repeated adamantly by cognitive scientists influenced by Heidegger—focused too much on abstraction. And this called for a change of perspective.

Now, certainly what the AI researchers understood from Heidegger's phenomenological description of the being-in-the-world structure was that the theoretical, objectifying, and epistemological stance reigning supremely in AI's received view (that is, GOFAI), deserved to be questioned in order to devise a new approach that incorporates coping-in-the-world in the whole picture. This is a plea for an ecological turn in cognitive science which is dismissive of mere coping-in-the-head attempting to bring the world as phenomenon to bear on cognitive science research. Therefore, this is also a plea for action instead of symbol-manipulating and mere world-representing without being involved in a situation. What is surprising though is that many investigations which adopted what Preston calls the 'Heideggerian alternative' (1993, p. 43) were developed at the MIT AI Laboratory, which clearly seems somewhat paradoxical. As Winograd asserts, "for those who have followed the history of artificial intelligence, it is ironic that this laboratory should become a cradle of 'Heideggerian AI'" (1995, p. 110). And he then adds that "it was at MIT that Dreyfus first formulated his critique, and for twenty years the intellectual atmosphere in the AI Lab was overtly hostile to recognizing the implications of what he said. Nevertheless, some of the work now being done at that laboratory seems to have been affected by Heidegger and Dreyfus" (*idem*).

In fact, AI researchers with Heideggerian leanings in theory and practice learned from some of the complexities already underscored by Dreyfus (1992). This led them to question the primacy of representation, the assumption that mental processes are quasi-linguistic or essentially sentential, and the surmise that the problem of the background was computationally tractable. This first Heideggerian sort of cognitive science was an interactionist alternative (Preston 1993, p. 51) which, nonetheless, was

not meant to decry *tout court* the possibility of machines displaying in the future cognitive prowess of some complex sort. Accordingly, no in-principle argument against AI was here at play. Thus, the Heideggerian alternative is no abandonment of cognitive science but a correction of some of the most problematic assumptions of cognitivism. The Heideggerian alternative accepts Heidegger's devastating critique of Cartesian philosophy and charges the received view of being a continuation of Cartesianism by technological means. So, far from a complete abjuration, the approach can be couched more accurately in terms of a critique of the received view in cognitive science premised on the internal world model assumption. Criticisms were mostly directed against the view that intelligence is achieved mainly by mental planning, that learning amounts to applying skills by means of a program of rules, and that understanding language means translating it into some sort of Mentalese from whence thoughts are represented. These points were already suggested in Dreyfus's critique of artificial reason. The point now was to put them to work by means of technological implementation on the basis of phenomenologically informed computer programs.

This implied to have a certain historical awareness of the field. According to the new researchers, a first sign that the received view needed an urgent change of perspective was computational intractability (also known as the problem of combinatorial explosion), which had always been pervasive in AI research. Such problem could no longer be tolerated and a fresh and less computational view that is able to explain anew what it takes for coping to take place, should be sought. Computational intractability arises when the attempt is made to solve every difficulty by mere computational force or exhaustive search through the problem space (Preston 1993). Exhaustive search, for example, is neither feasible for humans when playing chess, nor for computers regardless of larger memories and faster computer power. Indeed, a game with a few instructions manual can easily get out of hand and turned into the problem of considering a million possibilities when looking for what to do next. Computational intractable is, for that matter, any problem which no computer can calculate in a reasonable amount of time, that is, where the time for receiving a computational solution increases exponentially (for instance, when the answer-return, due to computational complexity, can be expected in 30,000 years). So in principle, a supercomputer created by a race of pandimensional beings to calculate the answer to the ultimate question of life, the universe and everything—like Deep Thought

in Douglas Adams's comedy science fiction series *The Hitchhiker's Guide to the Galaxy*—will also have to make do with the problem of computational intractability when there are simply too many possibilities to be considered. To be sure, Douglas Adam's Deep Thought is only capable of avoiding computational intractability because it is a product of fiction.

Having all these aforementioned problems in mind, Agre and Chapman (1988) argued that—when it comes to the application of computing to everyday life situations—planning is computationally intractable. According to them, the complexity inherent in planning concerns three aspects: (i) the problem of prediction in a world of uncertainty and change, (ii) the necessity of accommodating the simplicity of executives by specifying plans in impractical detail, and (iii) the larger unaddressed issue of relating plan texts to concrete situations in the world. For these researchers, everyday activity is improvisatory in nature, involving a continual redecision on what to do now. Hence, over against the plan-as-program view typical of GOFAI, Agre and Chapman propose a plan-as-communication view, whose contrast to the old view sounds strikingly Heideggerian (and it admittedly is):

> in the world of improvisation, one assumes that things are not likely to go according to plan. This is not to say that the resulting activity is chaotic in nature. It is, however, to say that the orderly nature of the activity, on whatever scale, does not arise from its having been mapped out ahead of time through the construction of a plan. Instead, the orderly nature of the activity arises through the interaction of an improvising agent and that agent's familiar world. (Agre & Chapman 1988, p. 20-21)

Unlike a renewed optimistic approach looking ahead for algorithmic solutions to computationally complex problems, what Agre and Chapman diagnose can be seen as a profound revision of presupposed assumptions which are to be abandoned if somehow computer science is to shed some light on human activity.

The purpose of the Heideggerian alternative is then more to extrapolate from Heidegger—and this to make do with inherited problems in AI—than to expound him or, for that matter, to remain in phenomenological philosophy. Accordingly, to overcome cognitivism does not mean to be done with cognitive science as a whole. Instead of dismissing cognitive science outright, the Heideggerian approach seeks to emphasize aspects of human experience which were given rather scanty consideration in the traditional view. Preston (1993) suggests that a link between the diagnostic approach to computational intractability and Heidegger can be devised via Marr's notion of

computational level. Marr's (2010)[89] computational investigation on vision as human representation of information processing, provided an analysis of routine everyday activity stressing the role of the environment and—what is even more interesting—suggesting that some specific environmental processes may not be themselves represented, although they may underwrite the achievements of cognition, catering to cognitive processes as their enabling context. Heidegger's phenomenological *Weltbegriff* can then be couched in terms of being the background of upper level know-how displayed in everyday activity. Dreyfus was, of course, the first to suggest that such Heideggerian appropriation could result in a devastating critique of AI's classical view. However, unlike Dreyfus's recalcitrant claim that cognitivism is doomed to failure, this new Heideggerian approach is certainly more sanguine than Dreyfus's critique of artificial reason. It attempts to show that what is needed is a change of the explanatory framework relying on internal representation and computation as the explanatory basis of behavior. The reasoning behind these expectations elaborated on a topic well glossed over by AI researchers: GOFAI's conceptual affinity with Cartesianism. So the name Descartes appeared in the discursive horizon of cognitive science and was first mooted by AI researchers inspired by Heidegger's critique of Descartes in the first division of *Sein und Zeit*.

Winograd was the first top-level AI researcher[90] to admit to working explicitly under the influence of Heidegger's phenomenology. According to his assessment, "the challenge of applying Heidegger's philosophy to computation is not just as a way of building more effective computer systems, but in understanding computers as part of the network of equipment within which we encounter our Being. Heidegger offers us much more than the opportunity to improve our programming techniques" (1995, p. 125). Winograd is also a revisionist of the received view and has offered, along with Flores, a detailed treatment of how the new foundations for computer design might benefit from incorporating into its discourse phenomenological concepts and insights (Winograd & Flores 1987).

---

[89] Marr's work was originally published in 1982.

[90] Winograd, Professor Emeritus of Computer Science at Stanford University, is director of the Human-Computer Interaction Group at Stanford and was founding member and president of the global organization Computer Professionals for Social Responsibility.

Drawing heavily from Dreyfus's Heidegger-interpretation, but also from Maturana's autopoetic understanding of an organism's relation to its environment and from Gadamer's hermeneutics, Winograd and Flores yielded the first steps for developing the theory of the new approach. Their theoretical path was explained as presenting a challenge to the assumptions of the rationalistic tradition based on previous laboratory work that emerged in the practice of computer technology, and this in order to offer some alternative directions for the design of computer-based tools (Winograd & Flores 1987, p. 8-9). The unchampioned rationalistic assumptions were basically the characterization of situations in terms of identifiable objects with well-defined properties, the generalization of rules to be applied to situations in terms of those objects and properties, and the logical application of rules to the situation of concern, drawing conclusions about what should be done (Winograd & Flores 1987, p. 15). At the same time, meaning was construed on the basis of a truth-theoretic characterization premised on both a system of rules by which sentences of a natural language could be translated into formulas of a formal language, preserving syntactically the essence of its semantics, and on another system of rules by which meanings and formulas in this formal language are determined in a systematic way by their meanings of their parts and the structure by which those parts are combined (Winograd & Flores 1987, p. 19). These two systems of rules are accompanied by a series of systematic rules of logic that explain the interrelation of the truth conditions for different formulas. The fundamental kind of sentence was the indicative which states that a certain proposition is true (*idem*). All these rationalistic assumptions were ultimately crowned by an information-processing psychology working as support for cognitive science, according to which all cognitive systems are symbol systems sharing a basic underlying set of symbol manipulating processes. A theory of cognition could then be couched as a program written in an appropriate symbolic formalism that, when implemented, would produce the behavioral traits devised by the program (Winograd & Flores 1987, p. 25).

On the basis of Heidegger's understanding of Being and Gadamer's notion of interpretation, Winograd and Flores intend to provide a set of new assumptions that can contribute to lay the grounds of a new foundation for computer design. These are a series of theoretical assumptions that pertain to a realm which is not in itself theoretical. In Winograd's and Flores's words, our implicit beliefs and assumptions cannot all be made explicit (1987, p. 32). The practices in terms of which our human

world is rendered intelligible cannot be made exhaustively explicit, lest we abstract them from its original source. The traditional program that this new approach seeks to amend has it that the background can be made explicit (and needs to be), so that specific rules making explicit routines and coping can be written in a program. However, the hermeneutic circle prohibits one from succeeding in such an attempt, because its recognition implies the impossibility of adopting a theoretical attitude about our background of practices, whose assumptions are always operating within the framework they provide. Recognizing the hermeneutic circle is also understanding understanding. On Heidegger's terms: "*if we see this circle as a vicious one and look out for ways of avoiding it, even if we just 'sense' it as an inevitable imperfection, then the act of understanding has been misunderstood from the ground up*" (SZ § 32, p. 153). The point is, adds Heidegger, "not to get out of the circle but to come into it in the right way" (*idem*). Winograd and Flores believe they can do this by eliminating representations, since our ability to act comes from our familiarity with the world and not from our knowledge of it, and by means of a reversal of the whole picture of cognition. Our access to the world takes place through practical involvement and bodily dexterity. If AI wants to succeed, it must implement programs taking into account this very access. All these insights in Winograd's and Flores's work are coupled with ideas coming from autpoietic biology (Maturana & Varela 1980), where learning is understood not as a process of the accumulation of representations of the environment but crucially as a continuous process of transformation of behavior through continuous change in the capacity of the nervous system to synthesize it.

Applications of this Heideggerian turn can also be found in human-computer interaction (Whiteside & Wixon 1988; Wixon & Wilson 1997) seeking to illuminate the context of usability of computers and how they constitute tools that can be used effectively. Given that much usability research has taken place within abstract contexts like laboratories, the question to be posed in this regard is how altering the context by construing it abstractly might also alter the meaning and the significance of the behavior observed. The opposite is actually what must be sought: "the detached analysis of the world in terms of abstract properties, which is a principal method of formalism and mechanism, is not primarily in Heidegger's scheme [on whom our work is explicitly based]. Instead, the starting point is ongoing experience, in the moment experienced, in the particular context experienced,

prior to explanation or analysis in terms of properties" (Whiteside & Wixon 1988, p. 371). Finally, work carried out by Flores and his coworkers in the 1980s (Flores 1981; Flores & Ludlow 1981; Flores & Graves 1986; Flores, Graves, Hartfield & Winograd 1988) explicitly sought to assimilate Heideggerian phenomenological concepts so that they could have a long-lasting bearing on the design of technology as a generative act that touches on the nature of human being. Although such purpose appears philosophical, the main idea was to use these notions of context and practice in the organizations of the corporative world. None other was the attempt (Spinosa, Flores & Dreyfus 1997) to render the ontological structure of everyday history-making usable for the skills of cultural innovation that characterize entrepreneurship.

Both Preston (1993) and Winograd (1995) suggest that the origins of the Heideggerian alternative can be traced back to the attempt to design programs based on ongoing activity, instead of writing plans premised on the surmise that some central system is responsible for constructing a map of the situation and for planning the paths through it. That there is no central processing 'headquarters' implies also that there is no central representation of the world, nor pictures in the head, nor a sentential language of thought. This new alternative view is fleshed out in practice in the work of MIT roboticist Rodney Brooks on behavioral-based robotics and of MIT trained AI researcher Philip Agre on the reorientation of the field away from thought and toward activity. Brooks is not admittedly Heideggerian, although his work has been influenced by researchers (Agre & Chapman 1988; Agre 1988) who have taken up the challenge of programming ready-to-hand (*Zuhandenheit*). So there it is: another way of characterizing the Heideggerian alternative in AI is to summarize the new approach as the conviction that the received view attempted (wrongly) to program metaphysical presence (*Vorhandenheit*), whereas it was actually *Zuhandenheit* or practical involvement what demanded engineering attention. According to the researchers enthused over the new possibilities provided by this change of perspective, what was needed was the design of programs that appropriately resemble the phenomenology of being-in-the-world. And this is the project, exemplified in Brook's and Agre's work, which shall be explained immediately.

Embodiment in Behavior-Based Robotics

Rodney Brooks is one of the world's most prestigious roboticists, if not the most renowned. Apart from having directed the Computer Science and Artificial Intelligence Laboratory at MIT, he founded the company iRobot, responsible for designing robots for NASA Mars exploration missions, and has been father of some famous artificial creatures like Genghis, Allen, Cog, and Baxter.[91] As was mentioned before, he denies explicitly a direct Heideggerian influence over his own work. As a matter of fact, in a famous essay titled 'Intelligence Without Representation,' he was quick to point out that his approach was not based on German philosophy or, for that matter, on Heidegger's work, although it "has certain similarities to work inspired by this German philosopher" (1999a, p. 97).

Industrial robot Baxter.

The importance of Brooks's work for the history of AI research resides in his being a pioneer in denying outright the dogma of cognitivism according to which cognition cannot be conceived without representation (Fodor 1975). Above all, Brooks mounted an overall critique of the traditional approach that emphasized the primacy of the abstract manipulation of symbols, "whose grounding in physical reality," on his own words, "has rarely been achieved" (1999a, p. 111). And he did this not only theoretically, but by showing how artificial creatures could be designed without subscribing to the aforementioned dogma. The new emphasis originating in Brooks's robots stresses the "ongoing

[91] Brooks has told every detail of the design of his creatures in his personal history of robotics (2003).

physical interaction with the environment as the primary source of constraint on the design of intelligent systems" (*idem*) and avoids handsomely some of the most recalcitrant paradoxes of the symbolic approaches (for example, the problem of describing with ridiculous detail in the guise of an universal encyclopedia of situations—like Lenat's Cyc [see Lenat, Prakash & Shepherd 1986]—the whole constitution of the background of activity in order to avoid any change paralyzing the program).

But as it happens in scientific communities, it is easier to find confirmations than refutations, and although Popper (2002) applied this pathology only to Marxists and Freudians (whom he despised), the divergent view being brought forth by Brook's turn toward behavior-based robotics, was not well received by the AI establishment when it was first proposed. This is perhaps a confirmation that (*pace* Popper) scientists are not that willing to look for refutations to their most beloved theories. It must be noted that even today the emphasis on situatedness and connectedness to the world is deemed too radical by computational cognitivists. On Bach's (2009) view, for instance, 'radical' behavior-based approaches cannot even render a satisfactory explanation as to why Stephen Hawking can actually interact with the world via a well-defined mechatronic interface: "in other words, tight sensor-coupling with a rich physical environment seems neither a sufficient nor a necessary condition for cognitive capabilities" (p. 25). Adams and Aizawa (2009) have argued that "for all that the radical philosophers have said, the mind is still in the head" (p. 78). Clark and Toribio (2001) even have referred to the sensimotor account of vision and visual consciousness (O'Reagan & Noë 2001), construed on the basis of the intrinsic coupling between the environment and enactive consciousness and perception, as 'sensimotor chauvinism,' arguing that there is no in-principle argument or *a priori* constraint pertaining the possibility of designing intelligence without environmental coupling.

Be that as it may, the new dogma of the new movement which computational cognitivists cannot stand was Brooks's proclamation that "the world is its own best model" (1999a, p. 115), which certainly looked like a direct appropriation of Dreyfus's idea that "the meaningful objects… among which we live are not a *model* of the world stored in our mind or brain; *they are the world itself*" (1992, pp. 265-266). From this starting point, Brooks disbelieves that the top-down disembodied and nonenviromentally-coupled approach can succeed in creating intelligent creatures, like the ones with which he intends to populate the solar system (Brooks & Flynn 1989). On the contrary, the approach

must seek for a way to design intelligence without representation or, as Brooks has also called it, intelligence without reason (1999a, p. 133 ff). The idea behind these assumptions, as unusual as they may sound, is that a center of intelligence is only imputed by the observer, since "the creature itself has none; it is a collection of competing behaviors" (1999a, p. 90). That might strike one as a Dennettian resort to the possibility of zombies: "are zombies possible? They're not just possible, they're actual. We're all zombies. Nobody is conscious—not in the systematically mysterious way that supports such doctrines as epiphenomenalism" (1991, p. 406). So, leaving aside that Dennett—unlike Brooks—conceives of the possibility of disembodied intelligent creatures, Brooks's robots might also be zombies in just this specified way. The need for a central system of operations might just be imputed by us, observers, from without.

Following Chapman and Agre (1988), Brooks hypothesizes that "much of even human level activity is similarly a reflection of the world through very simple mechanisms without detailed representations" (1999a, p. 91). On this account, even the early attempts at simplifying the world through the creation of microworlds has disastrous consequences. Unlike such caricaturization of the highly complex dynamics of the real world, which is anchored in the false idea that by impoverishing the world one might gain its richness piece by piece, for Brooks the point is to test creatures in the real world, that is, in the same world inhabited by humans. So in overt contradiction with the traditional idea that "intelligence is the work of symbol systems" (Simon 1996, p. 23), Brooks proposes to ground design in the world by means of a subsumption architecture enabling "to tightly connect perception to action, embedding robots concretely in the world" (1999a, p. 116). This subsumption architecture is the one enabling robots to use the world as their own model, continually referring to their sensors rather than to an internal world model (1999a, p. 166).

In words resembling the critique of rationalism in AI (Winograd & Flores 1987), Brooks is convinced that top down notions like *thought* and *reason* are metaphors of intelligence that followed naturally from the stored-program of von Neumann's model of computation, consisting of a processor (assumed to be functionally identical with the human brain) that executes instructions separated from a memory containing data and programs (Agrawala & Noh 1992). This 'von Neumanesque' computer architecture led AI in particular directions and dead-end alleys, so there was the temptation (often

succumbed to) of designing programs grounded on a cerebral and hierarchical notion of intelligence commencing from exhaustive planning, problem-solving, and previous representation modeling. However, for an interactionist approach that tightly connects perception with action, the question boils down to what people do in their everyday lives. The point is then to investigate this source of real prowess that can be found in human coping. As far as one can tell, people do not spend their lives reflecting about what to do next or planning about the different possibilities for action. And they seem to traverse the world ignoring environmental changes unimportant for doing the coping. Not even the use of objects in the world is preceded by mental representations or the purported semantic correspondence of these representations with symbols that the agent already possesses. On the contrary, objects can rather be defined through interactions of the agent with the world (Agre & Chapman 1987). Hence, object manipulation comes first. Reflection upon objects is just secondary.

Consequently, the style of work exhibited in Brooks's robot design (Brooks 1999a, pp. 138-139) can be characterized as showing four features:

·*Situatedness*: the robots are situated in the world, which means that they do not primarily deal with abstract descriptions or symbols needing subsequently to be related with the external environment. They deal with the here and now of the world directly influencing the behavior of the system.

·*Embodiment*: the robots have bodies and so experience the world directly (not just 'mentally'). For that matter, their actions are part of a dynamic with the world having immediate feedback on their sensations.

·*Intelligence*: intelligence can certainly be imputed on these robots but the source of their intelligence should not be limited to just the computational engine. Their intelligence comes particularly from their situatedness in the world, the signal transformations within their sensors, and the physical coupling taking place between them and the environmental dynamics.

· *Emergence*: intelligence can thus be said to emerge from the system's interactions with the world and also sometimes from indirect interactions between its components.

The traditional approach has tended to "amplify the abstraction away from situatedness, or connectedness to the world" (Brooks 1999a, p. 150). On Brooks's view, however, "without ongoing participation and perception of the world there is no meaning for an agent" (1999a, p. 168). Perception is participation in a situation, so it can hardly be affirmed that representing is original. Moreover, a human agent is embedded in her environment. This 'in,' as Dewey argued in his *Logic* (1938), is not to be taken as when we say that "pennies are 'in' a pocket or paint is 'in' a can… Interaction is going on between individuals and objects and other persons. The conceptions of situation and of interaction are

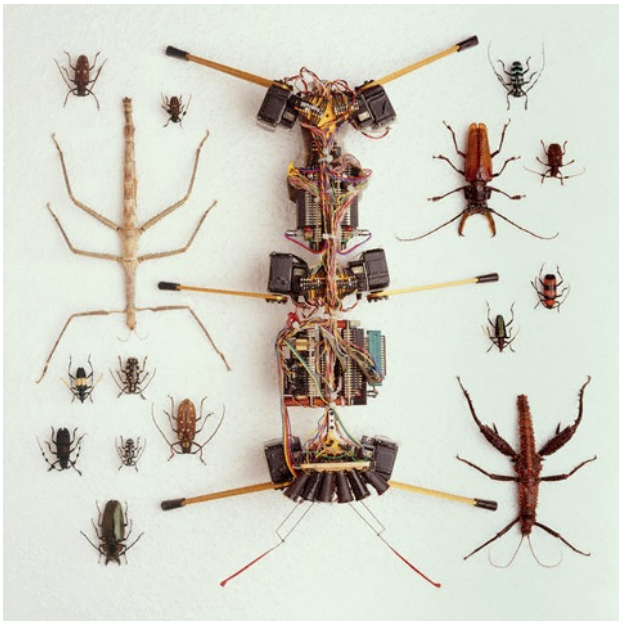
inseparable from each other" (quoted by Gallagher 2009, p. 39). On Brooks's view, the same pragmatic features must be applied to artificial agents if they are somehow to show the prowess and dexterity that can be deemed 'intelligent.'

From Brooks's approach—which, again, was not only a theoretical postulate but, more importantly, was backed-up in praxis by his successful artificial creatures—a direct threat to the customary biases of traditional AI might be deduced. Particularly, AI's tendency to view cognition as fully explainable by inputs and internal processes that could be broken into structure states and functional transformations, along with its reductionist epistemology according to which knowledge consists of enumerable discrete elements, its computational metaphor that skills were simply compiled from previously known facts and rules, and its persistence in the stored-program memory metaphor (Clancey 2009, pp. 14-15). A 'von Neumanesque' computer architecture implies a view of intelligent organization in terms of a hierarchy: there is an perceiver-pole, an 'I,' mapping external events in the world onto internal representations, which can then be used to plan action and executing activities. This is the so-called sense-think-act model (Pfeifer & Bongard 2007, p. 134) against which Brooks revolted in the early 1980s. Brooks's subsumption architecture can be interpreted also as viewing intelligence as the emergent result of parallel, asynchronous processes that are only loosely coupled over against the traditional view of hierarchically coupled processes. Pfeifer and Bongard elaborate on this contrast of loosely coupled processes and hierarchically coupled ones:

> In the latter there is a control program (the 'I') that calls the subroutines (e.g., for perception), and the calling program then has to wait for the subroutine (the perceptual act) to complete its task before if can continue (and go on to the action planning phase and then the action phase). This hierarchical control corresponds to very strong coupling; there is a very tight control regime between the calling and the called routines… 'Loosely coupled' also refers to the coupling of subsystems of an agent through its interaction with the environment… The coupling is called 'loose' because the global coordination is achieved indirectly—through the environment—and not directly through the neural system. (2007, p. 135)

As can be expected, hierarchical coupling systems tend to be rather rigid and cumbersome when it comes to performing tasks in the ambiguous, uncertain, dynamic, and open environment of the real world. It comes as no surprise that Brooks's preference is towards designing robots like Genghis, an artificial creature that can mimic the movement of insects and learns to walk from its mistakes.

Unlike Cartesian robots (which would be required to follow strictly the principles of operation and would be only functional in deterministic systems or petty microworlds), bug-bots, insect-inspired



Genghis and some autonomous arthropods.

robots, or autonomous arthropods are flexible, robust, energy efficient and, more importantly, *open* in their design manner of adapting to the operational environment. A 'von Neumanesque' hierarchical architecture supposes a CPU (central processing unit) with sequentially executed tasks organized in a complicated modular organization. In contrast, autonomous robots imply "no master-slave relationships among modules, no explicit functional or signal connections among them, no fixed hierarchical structure, and no center of control" (Gomi 1997, p. 98). Intelligence can thus be defined as emerging from the self-organization resulting from the dynamics of activity, and not as a one-to-one correspondence between cause and effect stimuli/action modules (Gomi 1997, p. 101).

In the preface to *How the Body Shapes the Way We Think* (2007), Pfeiffer and Bongard summarize research and technological implementation influenced by the embodiment turn:

> Research labs and leading technology companies around the world have produced or are developing a host of sometimes science fiction-like creatures: almost frighteningly realistic humanoid robots, robot musicians, wearable technology, robots controlled by biological brains, robots that can walk without a train, real-like cyborgs, robots in homes for the elderly, robots that literally put themselves together, artificial cells grown automatically, and simulated genetic regulatory networks for growing virtual creatures. This new breed of technology, along with many significant theoretical advances, is the direct result of the embodied approach to intelligence. (2007, p. xvii)

To be sure, Brook's work was path-breaking in the embodied approach and it was courageously proposed when it ran completely against the mainstream in the field (Pfeiffer & Bongard, p. 353). It is controversial, however, whether this artificial self-organizing systems can give rise to forms of more

complex intelligence, like the ones exhibited by human beings or even the higher animals. Brooks's failed attempt to jump from insect to human in the long-term Cog project, was mocked by Dreyfus: "of course, the long term project was short lived" (2007a, p. 250). Objections of like nature to Brooks's project will be resorted to in due course. But first, another approach, one which proudly recognizes itself as Heideggerian, shall be dealt with. Namely: Agre's shift from a mentalist to an interactionist conception of computation.

Making AI Philosophical Again

Philip Agre received his doctorate in Electrical Engineering and Computer Science at MIT, but he was always more interested (than many of his peers, that is) in exploring the philosophical assumptions pervading technological practices.[92] Thus, he deemed 'mistaken' to consider the Cartesian lineage of AI ideas as merely incidental and as having no purchase on the technical ideas that descend from it (1997, p. 23). Quite on the contrary, argues Agre, "computer systems are thus, among other things, also philosophical systems—specifically, mathematized philosophical systems—and much can be learned by treating them in the same way" (1997, p. 41). Agre claims that "AI is philosophy underneath" (2005, p. 155); an assertion he clarifies in five points:

> · AI ideas have their genealogical roots in philosophical ideas.
>
> · AI research programs attempt to work out and develop the philosophical systems they inherit.
>
> · AI research regularly encounters difficulties and impasses that derive from internal tensions in the underlying philosophical systems.
>
> · These difficulties and impasses should be embraced as particularly informative clues about the nature and consequences of the philosophical tensions that generate them.
>
> · Analysis of these clues must proceed outside the bounds of strictly technical research, but they can result in both new technical agendas and in revised understandings of technical research itself. (*idem*)

Influenced heavily by Dreyfus's pragmatization of Heidegger, Agre too understands *Sein und Zeit* as providing a phenomenology of ordinary routine activities, and believes Heidegger's *Daseinsanalytik* can

---

[92] In the informal essay, 'Critical Thinking for Technical People,' (originally an email message for the subscribers of the Red Rock Eater News Service—a popular electronic mail service organized by Agre that was active in the mid 1990s, credited as a model for many of today's political blogs and online newsletters), Agre tells the story "about how I became (relatively speaking, and in a small way) a better person through philosophy." See online: <http://polaris.gseis.ucla.edu/pagre/rre.html>.

provide useful guidance for the development of computational theories of interaction.[93] Most importantly, it can also contribute to afford technical practice a historical conscience it overtly lacks, since "research like Heidegger's can have its most productive influence upon AI when AI itself recovers a sense of its own historical development" (Agre 1996, p. 25). This last critical and historical trait permits Agre, as a philosopher of computing, to denounce that modern computational practices can be viewed as the resolute incarnation of a disembodied conception of philosophy having Augustine, Descartes, and Turing as pivotal figures, with the opposition of body and soul at the core of their thinking:

> Each man's cultural milieu provided fresh meaning for this opposition: Augustine struggled to maintain his ideals of Christian asceticism, Descartes described the soldier's soul overcoming his body's fear as the Thirty Years' War raged, and Turing idealized disembodied thought as he suffered homophobic oppression in modern England. (1997, p. 103)

This means, for Agre, that there is a historical tradition and discourse sustaining the practices of contemporary computational approaches, so by no means they can be said to sustain themselves exclusively on technical terms. The latter view is not only naïve but also dishonest. But unfortunately for the field, Agre sees that computer science is utterly oblivious to "its intellectual contingency and recast itself as pure technique" (*idem*). This is the reason why Agre castigates this forgetfulness of the assumptions running deep in AI, which more often than not are compensated for, put aside, and substituted by the formalist attempt to cleanse computational programs of the 'inexactness' of natural language, and to strip AI altogether of its historical and cultural underpinnings. It is by virtue of not paying attention to how their scientific practices are constituted that formalists attempt to liberate computational work precisely from the unruliness and imprecision of vernacular language, which appears foreign and annoying to their technical field. Moreover, "they believed that, by defining their vocabulary in rigorous mathematical terms, they could leave behind the network of assumptions and associations that might have attached to their words through the sedimentation of intellectual history" (Agre 2002, p. 131). This is why Agre believes such an attempt should not be countenanced any longer but rather it should be confronted by means of a 'critical technical practice:' the kind of critical stance that would guide itself by a continually unfolding awareness of its own workings as a

[93] For a thorough 'phenomenology of everyday life,' framed along the lines of a pragmatization of Heidegger, see Pollio, Henley & Thompson (1997).

historically specific practice (1997, p. 22). As such, "it would accept that this reflexive inquiry places all of its concepts and methods at risk. And it would regard this risk positively, not as a threat to rationality but as a promise of a better way of doing things" (Agre 1997, p. 23).

This critical technical practice proposed by Agre has clear overtones oscillating amid an immanent critique, on the one hand, and an 'epistemological electroshock therapy' toward situating scientific knowledge, on the other (Haraway 1988). Agre's view, then, might appear problematic if it is ill-construed as confusing or ambiguous, since it can be seen both as a critique *from within*—accepting the basic methodology and truth-claims of computer science, peppered with internal disputes against the more obviously invalid and politically loaded claims—and as a critique *from without*—recognizing the ultimate cultural contingency of all claims to scientific truth (Sengers 1995, p. 151). Nevertheless, it is crucial for Agre to present his work as neither an internalist account *of* AI, nor as a philosophical study *about* AI, but as "actually a work *of* AI: an intervention within the field that contests many of its basic ideas while remaining fundamentally sympathetic to computational modeling as a way of knowing" (1997, p. xiv). As was the case with Brook's work on robotics, Agre also finds it more daring to intervene in the field and show practically how critical technical views might help develop better artificial systems. So both the critical intervention on the field and the fundamental sympathetic posture deserve, furthermore, a separate explanation.

Concerning the critical intervention, Agre notes that there is a certain mindset when it comes to what 'computer people' believe—and this is, of course, Agre's niche—regarding the aims and scope of their own work. This belief is not at all arbitrary or merely capricious, but rather it must be viewed in conjunction with the very nature of computation and computational research in general. According to Agre, computational research can be defined as an inquiry into physical realization as such. Moreover, "what truly founds computational work is the practitioner's evolving sense of what can be built and what cannot" (1997, p. 11). The motto of computational practitioners is simple: *if you cannot build it, you do not understand it. It must be built and we must accordingly understand the constituting mechanisms underlying its workings*. This is why, on Agre's account, computer scientists "mistrust anything unless they can nail down all four corners of it; they would, by and large, rather get it precise and wrong than vague and right" (1997, p. 13). So there is also a 'work ethic' attached to this computationalist mindset: *it has to*

work. However, Agre deems it too narrow to entertain just this sense of 'work.' Such conception of what counts as success is also ahistorical in that it can simply be defined as working because the program conforms to a pre-given formal-mathematical specification. But an AI system can also be said to work in a wholly different sense: when its operational workings can be narrated in intentional terms by means of words whose meanings go beyond the mathematical structures (which is, of course, a pervasive practice in cognitive scientific explanations of mechanism). For example, when a robot is said to 'understand' a series of tasks, or when it is proclaimed that AI systems will give us a deeper insights about human thinking processes. This is indeed a much broader sense of 'work,' one which is not just mathematical in nature, but rather a clearly discursive construction. And it certainly bears reminding that such discursive construction is part of the most basic explanatory desires of cognitive science. So in the true sense of the words 'build' and 'work,' AI is not only there to build things that *merely* work. Let us quote at length:

> The point, in any case, is that the practical reality with which AI people struggle in their work is not just 'the world,' considered as something objective and external to the research. It is much more complicated than this, a hybrid of physical reality and discursive construction. The trajectory of AI research can be shaped by the limitations of the physical world—the speed of light, the three dimensions of space, cosmic rays that disrupt memory chips—and it can also be shaped by the limitations of the discursive world—the available stock of vocabulary, metaphors, and narrative conventions. (Agre 1997, p. 15)

This also gives hints as to how exogenous discourses, like philosophy, are supposed to be incorporated into technological practices. Agre is of the opinion that the point is not "to invoke Heideggerian philosophy, for example, as an exogenous authority that supplants technical methods. (This was not Dreyfus's intention either.) The point, instead, is to expand technical practice in such a way that the relevance of philosophical critique becomes evident *as a technical matter*. The technical and critical modes of research should come together in this newly expanded form of critical technical consciousness" (1997, p. xiii). The critical technical practice Agre envisions is one "within which such reflection on language and history, ideas and institutions, is part and parcel of technical work itself" (2002, p. 131). More exactly, Agre confesses that his intention is "to do science, or at least something about human nature, and not to solve industrial problems" (1997, p. 17). And he adds: "but I would also like to benefit from the powerful modes of reasoning that go into an engineering design rationale" (*idem*). In such a way, Agre pretends to salvage the most encompassing claims of AI research

—that it can teach us something about the world and about ourselves—by means of incorporating a self-correcting, history-laden approach combining both technical precision and philosophical rigor. By expanding the comprehension of the ways in which a system can work, "AI can perhaps become a means of listening to reality and learning from it" (Agre 2002, p. 141). But it is precisely because of its not having listened to reality that, for instance, Dreyfus launched his attacks against AI as an intellectual enterprise.

Agre contends that merely "lashing a bit of metaphor to a bit of mathematics and embodying them both in computational machinery" (1997, p. 30)—which is usually what computer scientists come up with—will not do the job of contributing to the understanding of humans and their world. So framed, the approach appears to Agre as too narrow, naïve, and a clear way of not listening to reality. So he has a more ambitious project: the very metaphors being lashed to a bit of mathematics that end up in machinery implementation must be investigated. Both physical reality and discursive construction must be taken into account. Although technical languages encode a cultural project of their own (the systematic redescription of human and natural phenomena within the limited repertoire of technical schemata that facilitate rational control)—a fact which tends to be as such elided—"it is precisely this phenomenon that makes it especially important to investigate the role of metaphors in technical practice" (Agre 1997, p. 34). At this juncture, Agre sounds strikingly similar to Blumenberg, whose metaphorological project "seeks to burrow down to the substructure of thought, the underground, the nutrient solution of systematic crystallizations; but it also aims to show with what 'courage' the mind preempts itself in its images, and how its history is projected in the courage of its conjectures" (2010, p. 5). For Agre too metaphors play a role in organizing scientific inquiry or, to say it with Blumenbergian tones, metaphors are by no means 'leftover elements' (*Restbestände*) but indeed 'foundational elements' (*Grundbestände*) of scientific discourse.[94] Clinging to Kuhnian terminology, this can also be couched in terms of the tension between normal science—with its aseptic attitude toward reducing instability of meaning and inconsistency via a cleansing of elements of inexact, ambiguous nature—and revolutionary science which makes metaphoric leaps that create new meanings and

[94] It must be noted that, for the sake of the argument, reference is here made to early Blumenberg and his project in *Paradigmen zu einer Metaphorologie* (1960) of tracing absolute metaphors as those *Grundbestände* that cannot be conceptually reduced (*nicht in Begrifflichkeit aufgelöst werden können*) but rather function as constituting a catalytic sphere from which the universe of concepts continually renews itself.

applications that might constitute genuine theoretical progress (Arbib & Hesse 1987, p. 157). By showing how technical practice is not only the result of technical work but also of discursive construction and unexplained metaphors, Agre's critical technical practice might meet the criteria for being considered a truly revolutionary approach in Kuhnian terms. It remains to be seen, however, whether that is indeed the case.

The sympathetic attitude towards computational modeling that Agre espouses takes as its point of departure the analysis of agent/environment interactions which accordingly should be extended to include the conventions and invariants maintained by agents throughout their activity. This notion of environment is referred to as *lifeworld* and can be incorporated into computational modeling via "a set of formal tools for describing structures of lifeworlds and the ways in which they computationally simplify activity" (Agre & Horswill 1997, p. 111). From this follows that, if *embodiment* was the central notion with which Brooks framed his robotics project, Agre's emphasis lies on *embedding*. The distinction between embodiment and embedding can be explained as follows:

> 'Embodiment' pertains to an agent's life as a body: the finiteness of its resources, its limited perspective on the world, the indexicality of its perceptions, its physical locality, its motility, and so on. 'Embedding' pertains to the agent's structural relationship to its world: its habitual paths, its customary practices and how they fit in with the shapes and workings of things, its connections to other agents, its position in a set of roles or a hierarchy, and so forth. The concept of embedding, then, extends from more concrete kinds of locatedness in the world (places, things, actions) to more abstract kinds of location (within social systems, ecosystems, cultures, and so on). Embodiment and embedding are obviously interrelated, and they each have powerful consequences both for agents' direct dealings with other agents and for their solidarity activities in the physical world. (Agre & Horswill 1997, 111-112)

The importance for cognitive science of having a well-developed concept of the environment is not to be underestimated, since it seems that only prior a basic understanding of an agent's environment can a given pattern of adaptive behavior be figured out. Taking a stride towards defining the environment with at least a modicum of rigor amounts to developing "a *positive theory* of the environment, that is, some kind of principled characterization of those structures or dynamics or other attributes of the environment *in virtue of which* adaptive behavior is adaptive" (Agre & Horswill 1997, p. 113). Accordingly, Agre and Horswill lament that AI has downplayed the distinction between agent and environment by fatally reducing the latter to a discrete series of choices in the course of solving a problem. On their view, "this is clearly a good way of modeling tasks such as logical theorem-proving and chess, in which the objects being manipulated are purely formal" (*idem*). AI can go on well without

a well-developed concept of the environment but only at the price of focusing on mere toy-problems, microworlds, and toy-tasks within such artificial environments. It should then not come as a surprise that the situation changes dramatically for tasks involving physical activities, where "the world shows up, so to speak, phenomenologically: in terms of the differences that make a difference for *this* agent, given its particular representations, actions, and goals." (*idem*). Such environmental indexicality which is brought forward here finds often the objections of cognitivists who criticize the view that agents perform tasks without any computation whatsoever, as though agents inhabiting a lifeworld lived in an adamant reactive mode. But the point is rather that "the nontrivial cognition that people do perform takes place against a very considerable background of familiar and generally reliable dynamic structure" (Agre & Horswill 1997, p. 118). Now, precisely indexicality has been difficult to accommodate within AI research. With this in view, Agre has criticized the usual assumptions of the received view as follows:

> · That perception is a kind of reverse optics building a mental model of the world by working backward from sense-impressions, inferring what in the world might have produced them.
>
> · That action is conducted through the execution of mental constructs called plans, understood as computer programs.
>
> · And finally, that knowledge consists in a model of the world, formalized in terms of the Platonic theory analysis of meaning in the tradition of Frege and Tarski. (2002, p. 132)

The dissociation of mind and body (the founding metaphor of cognitive science and modern philosophy) is here at work, precisely when traditional AI thinks of the mind roughly as a plan generator and the body as the executor of the plan. Moreover, AI is so framed in terms of a series of dissociations: mind versus world, mental activity versus perception, plans versus behavior, the mind versus the body, and abstract ideas versus concrete things (Agre 2002, p. 132). According to Agre, these dissociations are contingent and can be considered 'inscription errors' (Smith 1996): "inscribing one's discourse into an artifact and then turning around and 'discovering' it there" (Agre 2002, p. 130). And this is not to be admired. As Nietzsche contented in *Über Wahrheit und Lüge im außermoralischen Sinne* (1873), when someone hides something behind a bush and looks for it again in the same place and finds it there as well, there is not much to praise in such seeking and finding.

That AI research has been framed along these contingent oppositions makes it clear that it is part of the history of Western thought. As such,

> it has inherited certain discourses from that history about matters such as mind and world, and it has inscribed those discourses in computing machinery. The whole point of this kind of technical model-building is conceptual clarification and empirical evaluation, and yet AI has failed either to clarify or to evaluate the concepts it has inherited. Quite the contrary, by attempting to transcend the historicity of its inherited language, it has blunted its own awareness of the internal tensions that this language contains. The tensions have gone underground, emerging through substantive assumptions, linguistic ambiguities, theoretical equivocations, technical impasses, and ontological confusions. (Agre 2002, p. 141)

Nevertheless, it is interesting to note that—for all his philosophical acumen—Agre himself has not been able to liberate himself from the persistence of a representational theory of cognition, even when his is certainly more concrete, more historically conscious, and more enactive than the one customarily held in the traditional view. As a result, the latter critical concepts grouped together conform the motivation for developing a concept of indexical-functional or deictic representation (Agre & Chapman 1987; Agre 1997), the main idea being that agents represent objects in generic ways through relationships to them (Agre & Horswill 1997, p. 118). On Agre's view, what must be done is refine the concept of representation and not cast it aside and show what kind of representational activity is at work in interaction. Thus, the point is to criticize the underlying view of knowledge presupposed by the traditional theory of representation (that knowledge is picture, copy, reflection, linguistic translation, or physical simulacrum of the world), while suggesting that "the primordial forms of representation are best understood as facets of particular time-extended patterns of interaction with the physical and social world" (Agre 1997, p. 222). Therefore, "the notion of representation must undergo painful surgery to be of continued use" (Agre 1997, p. 250). Given that this redefinition of representation by Agre has its own quirks, it must now be carefully explained.

The traditional theory of representation which has been put into work and is thoroughly presupposed in AI research is based on the notion of world model. Such notion refers to some structure which is thought to be within the mind or machine that represents the outside world by standing in a systematic correspondence with it (Agre 1997, p. 223). As such, the assumption that there is a world model being represented by the mind is the epitome of mentalism (Agre 1997, p. 225). Mentalism was previously defined by Agre as the generative metaphor pervasive in cognitive science according to which every human being has an abstract inner space called a 'mind' which clusters around a dichotomy between outside and inside organizing a special understanding of human existence (Agre 1997, p. 49). Marres (1989), a defender of mentalism, defines it as the view that the mind directs the

body. Thus, on Agre's terms, giving preeminence to indexicality amounts to inverting this picture, since conceding that human beings are not minds that control bodies implies that interaction cannot be defined "in terms of the relationships among a mind, a body, and an outside world" (1997, p. 234), which is unfortunately so typical in cognitive scientific explanations. And here the key term is indeed interaction, understood not as the relation between the subjective and the objective, but rather as emerging from the actual practices people employ to achieve reference *in situ*. So indexicality "begins to emerge not merely as a passive phenomenon of context dependence but as an active phenomenon of context constitution" (Agre 1997, p. 233).

Chalmers could only table the question *what is it like to be a thermostat?* (1996, p. 293) by means of importing some heavy philosophical baggage, namely the assumption that the thermostat controls the temperature of systems in general (not of *this* specific system, say the internal combustion engine of a specific car), or that a thermometer measures the temperature "in room 11" (instead of *here*), or that one eats with "fork number 847280380" in some cosmic registry of forks (instead of *precisely this* fork I am holding with my left hand). Quite on the contrary, when indexicality is introduced as a constituting factor of interaction, it turns out that "human activities must be described in intentional terms, as being *about* things and *toward* things, and not as meaningless displacements of matter. Physical and intentional description are not incomparable, but they *are* incommensurable" (Agre 1997, p. 245). From this follows that Chalmers's ascription of intentional states to nonembedded systems is absurd, precisely because embedding, and the interaction deriving thereof, is the condition of possibility for intentional comportment to take place. The ubiquitous character of experience suggested by Chalmers is also an inscription error, for it arises from obviating the need for a proper theory of intentionality or, to be more exact, such view derives from the naturalization of intentionality. When actual, concrete intentional activities are taken off the picture, representation is no longer connected with a lifeworld. Thus, the illusion can be then entertained that a semantic theory merely entails the categorization in some objective way of the ontology of a concrete situation, before the event of activity has taken place, or ignoring *tout court* the eventual character of activity (Agre 1997, p. 232).

Incidentally, the aforementioned illusion is Agre's critique of the semantic theory espoused by Barwise and Perry (1983), which, on Agre's criticism, comports a metaphysical realism that obscures

indexicality. According to Agre, "when a speaker uses an indexical term such as 'I,' 'you,' 'here,' 'there,' 'now,' or 'then' to pick out a specific referent, this picking out is determined by relations between situations; it is not an *act* on the speaker" (1997, p. 233). So these interactions and how they shape situations must be clarified, since it can be said that "interaction is central, both to human life and to the life of any agent of any great complexity" (Agre 1997, p. 234). Embedded activities must be investigated in how they are structured, as well as the sort of representing which is most incumbent on them.

For Agre, the latter requires a proper theory of intentionality couched within the Heideggerian distinction between *Zuhandenheit* and *Vorhandenheit*. AI research can be accused of having only paid attention to present-at-hand phenomena, thus attempting to model computationally what precisely appears salient objectively in perception. In contrast, Agre finds that this Heideggerian distinction is not psychological nor mechanistic but a description of the structure of everyday experience which can be suitable for a new way of computational modeling of that experience. Preston (1993) had already explored this Heideggerian distinction in relation to another one: that of nonrepresentational and representational intentionality. One could, *à la* Dreyfus, identify respectively *Vorhandenheit* with representational intentionality and *Zuhandenheit* with a sort of nonrepresentational intentionality and so proclaim beforehand the failure of artificial systems propounding the accomplishment of high-level intelligence. For Agre, however, this is too radical and, above all, too pessimistic. What is needed is a clarification of what kinds of representation exist and the role they play in real activities (Agre 1997, p. 237). Herein resides the importance of delving into experience and providing AI with a set of tools to enrich its vocabulary and metaphors. This is needed because "the philosophy that informs AI research has a distinctly impoverished phenomenological vocabulary, going no further than to distinguish between conscious and unconscious mental states" (Agre 1997, p. 239). Agre is onto something more important here, which is nothing less than making AI philosophical again: "technology at present is covert philosophy; the point is to make it openly philosophical" (1997, p. 240).

The traditional idea of representation understood it as a model in an agent's mind that corresponds to the outside world through a systematic mapping. Agre opines that AI research has been concerned with only a partly articulated view of representation. No wonder, then, the meaning of

representations for an agent can be determined almost as *en-soi*—to use Sartre's terminology in *L'être et le néant* (1943)—without any reference being provided as to the agent's location, attitudes, interests, and idiosyncratic perspective (as *être-pour-soi*). This is also the reason explaining why "indexicality has been almost entirely absent from AI research" (Agre 1997, p. 241). Moreover, "the model-theoretic understanding of representational semantics has made it unclear how we might understand the concrete relationships between a representation-owning agent and the environment in which it conducts its activities" (*idem*). On Agre's view, the reason why AI research has lagged behind a clear-cut understanding of representation and indexicality has not been its nondistinctiveness between mechanism and human phenomena. Notwithstanding Agre's crucial imports from the alien province of phenomenology, he would nevertheless defer to Chalmers's highly controversial idea that experience is ubiquitous, albeit with a caveat: the problem is not to ask whether there is something it is like for a thermostat to be what it is, for Agre has it that any device that engages in any sort of interaction with its environment can be said to exhibit some kind of indexicality (1997, p. 241). Chalmers's problem is simply not to have considered exactly which kind of intentionality might be ascribed to artifacts like thermostats. Artifacts do have some sort of ambience embedding. So for instance "a thermometer's reading does not indicate abstractly 'the temperature,' since it is the temperature *somewhere*, nor does it indicate concretely 'the temperature in room 11,' since if we moved it to room 23 it would soon indicate the temperature in room 23 instead. Instead, we need to understand the thermometer as indicating 'the temperature *here*'—regardless of whether the thermometer's designers thought in those terms" (*idem*). As Agre's contention goes, the point is to ascribe indexicality to artifacts. In fact, "AI research needs an account of intentionality that affords clear thinking about the ways in which artifacts can be involved in concrete activities in the world" (1997, p. 242).

Such account of intentionality was coined by Agre under the rubric of deictic representation as opposed to objective representation. First, two sorts of ontology are to be distinguished. According to an objective ontology, individuals can be defined without reference to activity or intentional states. A deictic ontology, by contrast, can be defined only in indexical and functional terms and in relation to an agent's location, social position, current goals and interests, and idiosyncratic perspective (Agre 1997, p. 243). Entities entering the space of whatever interaction with the agent can only be understood

correctly in terms of the roles they play in the agent's activities. In accordance with the aforementioned deictic notation introduced by Agre, "some examples of deictic entities are *the-door-I-am-opening*, *the-stop-light-I-am-approaching*, *the-envelop-I-am-opening*, and *the-page-I-am-turning*. Each of these entities is indexical because it plays a specific role in some activity I am engaged in; they are not objective, because they refer to different doors, stop lights, envelopes, and pages on different occasions" (*idem*). Their nonobjective character, however, does not imply that indexical entities are to be considered, by contrast, subjective and, for that matter, phantasms or internal and intimate qualia. The idea behind this is precisely that a deictic ontology should not be confused with subjective, arbitrary musings of an encapsulated subject. In the first place, this is the ontology which can be most properly ascribed to routine activities. So it would be preposterous to suggest that they are intimate or ineffable. Routines activities are realized 'out there' in the world and, for that very reason, do not pertain to an internal mental game: they are, indeed, public. Accordingly, in routine activities the objective character of entities with which one copes, is not salient or important. Neither is their 'subjective feel,' or the way they appear to me as individual. That their character is deictic means that what is most important is the role they play in the whole of activity. Therefore, hyphenated noun phrases like *the-car-I-am-passing* or *the-coffee-mug-I-am-drinking-with* are not mental symbols in the cognitivist sense. They designate "not a particular object in the world, but rather a role that an object might play in a certain time-extended pattern of interaction between an agent and its environment" (Agre 1997, p. 251).

Agre's alternative way of conceiving of activity and the express purpose of modeling it computationally is very attractive. As a matter of engineering, the leading principle is that of machinery parsimony: "choosing the simplest machinery that is consistent with known dynamics" (Agre 1997, p. 246). This view explicitly contrasts with the emphasis on expressive and explicit representation typical of traditional AI, with all the inherent difficulties of programming beforehand, as scripts, all the situations an artificial agent might encounter when coping with the world. By clear contrast with GOFAI, "the principle of machinery parsimony suggests endowing agents with the minimum of knowledge required to account for the dynamics of its activity" (Agre 1997, p. 249). In such a way, Agre's approach also resonates with Brooksian tones of removing 'intelligence' and even 'reason' from the picture in order to render an account of interactive representation. Moreover, Agre sees deictic

representation as changing the traditional view altogether since it presents us with the possibility, not of expressing explicitly and in every detail objective states of affairs, but of *participating* in them: "conventional AI ideas about representation presuppose that the purpose of representation is to express something, but this is not what a deictic representation does. Instead, a deictic representation underwrites a mode of relationship with things and only makes sense in connection with activity involving those things" (1997, p. 253). However, the objection may be raised that such a deictic approach violates the grand spirit of AI which seeks greater explicitness of representation and broader generality, whereas Agre's formula for design might simply contribute to model only special-purpose—and thusly limited—devices. But Agre respondes that "the conventional conception of general-purpose functionality is misguided: the kind of generality projected by current AI practice (representation as expression, thought as search, planning as simulation) simply cannot be realized" (1997, pp. 249-250).

This is, of course, not just a series of theoretical postulates urged by Agre, since he distinguishes amongst levels of analysis (1997, pp. 27-28). The *reflexive* level, which has been already exhibited in the last pages of this exposition, provides ways for analyzing the discourses and practices of technical work. Given that technical language is unavoidably metaphorical, the reflexive level permits one to let those metaphors come to the surface and thus can they be taken into account when technical work encounters trouble in implementation. On the *substantive* level, the analysis is carried out with reference to a particular technical discipline, in this case AI. But Agre is primarily interested in proceeding, on top of the reflexive and substantive levels, on a *technical* level, in order to explore "particular technical models employing a reflexive awareness of one's substantive commitments to attend to particular reality as it becomes manifest in the evolving technical work" (1997, p. 28). On Agre's view, this partitioning of levels of analysis has not been conscientiously attended to by traditional AI practitioners. Particularly, the reflexive level that prescribes an awareness of the role of metaphors in technical work has been disdained, as though AI researchers could simply bootstrap their way to technical success without being aware of the underlying metaphors pervading their work. For Agre, this is specially problematic because "as long as an underlying metaphor system goes unrecognized, all manifestations of trouble in technical work will be interpreted as technical difficulties

and not as symptoms of a deeper, substantive problem" (1997, p. 260). As historical proof, just recall[95] how Papert (1968) reacted with outrage at Dreyfus's suggestions that GOFAI's problems were not merely 'technical,' but worse, substantive.

As an exemplary case of technical work based on the aforementioned levels of analysis, Agre presents Pengi, a program designed by Chapman and Agre (1987) in the late 1980s under the rubric of being an implementation of a theory of activity. Pengi is a penguin portrayed in the commercial computer game Pengo, who finds itself in a maze made up of ice blocks that is surrounded by an electric fence. The maze is also inhabited by deadly bees that are to be avoided at all costs by Pengi and the task of the player is to maintain Pengi alive and defend it from such perils coming along the way. As defense, the bees can be killed by crushing them with a moving ice block or by kicking the fence while they are touching it. This momentarily stuns the bees and they can be crushed by simply walking over them. Agre agues that Pengo is an improvement on the blocks world, although it obviously fails to capture numerous elements of human activity. What is important is the combination of goal-directedness and improvisation involved in the game, from which Agre hopes to learn some computational lessons. First of all, Agre and Chapman did not attempt to implement in advance everything they knew about the game, thus contradicting the mapping out beforehand which is typical in traditional AI systems. The point is to see Pengi as relating to the objects that appear in its world, not in terms of their resemblance to mental models which were beforehand programmed, but solely in terms of the roles they play in the ongoing activity. As such, what Agre and Chapman attempted to program was actually deictic representations: *the-ice-cube-which-the-penguin-I-am-controlling-is-kicking*, *the-bee-I-am-attacking*, *the-bee-on-the-other-side-of-this-ice-cube-next-to-me*, etc.

At any rate, Agre does not argue that this simple system can be regarded as intelligent: "Pengi does not understand what it is doing. No computer has ever understood to any significant degree what it was doing" (1997, p. 301). But the bottom line is straightforward enough to explain: the game constituting Pengi's world as agent is not made up of present-at-hand entities and processes, but more importantly of possibilities for action that require appropriate responses from the agent. This shows Agre's understanding of ready-to-hand entities as no entities at all, but as possibilities for action and

[95] This episode was resorted in Chapter 2.

subsequent responses to the demands of the situation at hand. Given that these possibilities for action are not objects at all and that usually this sort of open stance for responding skillfully to environmental challenges does not appear in propositional referring, it is understandable that they have been rather elusive for programers. After all, how can one program possibilities for action, since the focus is not on this particular object or the other but rather on the movement constituting the towards-which for-the-sake-of-which? The wellspring of this movement is all the more elided because, as Heidegger has it, precisely what is closest to us ontically is ontologically (and for that very reason) that which is farthest (SZ § 5, p. 15). This has been Agre's task, namely: to attempt to reveal the ontological dimension by means of a specific technological implementation that does not obfuscate it but rather embrace it. By programming deictic representations instead of just objective ones, Agre argues, computational programs can learn this fundamental lesson: what was lacking in GOFAI systems was precisely a model to envision a specific relationship between machinery and dynamics based on the concept of interaction. This lesson, so the argument goes, can gradually dispel the need for mentalist approaches. With this attempt, Agre has tried to program *Zuhandenheit* instead of *Vorhandenheit*. That this can be made is, however, highly controversial. Certainly, what is deeply contentious is not that phenomenological insights can be brought to bear on cognitive science for a critical technical practice like the one Agre requires, but rather the assumption that the experiential dimension which phenomenology has revealed can be programmable. According to Heidegger, "*the essence of Dasein lies in its existence*" (SZ § 9, p. 42), which does not imply any "'properties' present-at-hand of some entity which 'looks' so and so and is itself present-at-hand" (*idem*). To exist as Dasein, then, implies that one's own existence has to be partly constructed, for existence is not already given and therefore is no program that can be run by any kind of hardware (Capurro 2004). So the Heideggerian *Sichöffnende und Offene* is perhaps not amenable to programing. To say it with Heideggerian overtones: programing can only be ontic but not ontological, since if the ontological were susceptible to programing, it would not be ontological. But before mounting a scathing critique of this attempt, another more recent appropriation of Heideggerian philosophy within cognitive science must be presented, namely Wheeler's, to which we turn to in the next chapter for scaling up the scope of Heidegger's reception in contemporary cognitive science.

| 9 |

NATURALIZING HEIDEGGER

Neo-Heideggerian Cognitive Science

Michael Wheeler's project for reconstructing the cognitive world (2005) can be viewed as a reflection on the philosophical foundations of cognitive science, concerned with helping in the search for a sort of Kuhnian revolution in the field (2005, p. 15). A project which is essentially Heideggerian. According to Wheeler's understanding of the field's history, this revolutionary twist has been emerging over the last few years as a response to orthodox cognitive science—basically, GOFAI and connectionism, that is, "most cognitive science as we know it" (*idem*). Although the countermovement has adopted various names throughout its brief existence[96] and despite its identity being admittedly somewhat amorphous, it is customary to refer to it as embodied-embedded cognitive science. And this, believes Wheeler, because embodiment and embedding are part of "a central and distinctive theoretical tendency within the more nebulous movement" (2005, p. 11). As Clark claims, "talk of mind as *intimately* embodied and *profoundly* environmentally embedded shimmers at the cusp of the cognitive scientific zeitgeist" (2012, p. 275). In order to pin down why this new science of mind (Rowlands 2010) abjures of orthodox cognitive science, its aims and scope need some clarification.

According to Wheeler, "the embodied-embedded approach revolves around the thought that cognitive science needs to put cognition back in the brain, the brain back in the body, and the body back in the world" (*idem*). Incidentally, this purpose is substantially akin to the ambitions of Andy Clark's 1997 book titled with a decisively Heideggerian connotation: *Being There. Putting Brain, Body, and World Together Again.* On Clark's understanding, the new trend thinks it necessary "to abandon the idea (common since Descartes) of the mental as a realm distinct from the realm of the body; to abandon the idea of neat dividing lines between perception, cognition, and action; to abandon the idea of an

[96] For Marsh (2007), for example, Wheeler's theoretical heroes fall within a loose coalition of anti-representationalism or anti-Cartesianism under the rubric of dynamical-, embodied-, extended-, distributed-, and situated theories of cognition, or DEEDS, to use an acronym. But his is hardly the only acronym in use. The movement is sometimes referred to as 4E cognitive science, standing for embodied, embedded, extended and enactive cognitive science (Rowlands 2010b), or as 4EA, adding affective cognition to the equation (Protevi 2010).

executive center where the brain carries out high-level reasoning; and most of all, to abandon research methods that artificially divorce thought from embodied action-taking" (1997, pp. xii-xiii). Wheeler adheres overtly to the principles encompassing this novel program. His own project, construed as a reflection on the philosophical foundations of cognitive science, targets precisely Cartesian philosophy as the mindset dominating cognitive science from which the new approach needs to escape (1995; 2008). As is widely known, Heidegger criticizes Descartes boldly in *Sein und Zeit* while claiming at the same time that the *cogito sum* is no firm footing—as Descartes supposed. On the contrary, claims Heidegger, "what he left undetermined when he began in this 'radical' way was the kind of Being which belongs to the *res cogitans*, or—more precisely—the *meaning of the Being of the sum*" (SZ § 6, p. 24). Bluntly put, on Heidegger's view, the Cartesian *cogito sum* gives us no special insight into the nature of the *sum* itself. Be that as it may, Wheeler wants to revise both the traditional interpretation to which Descartes has often been subjected to in Anglo-American philosophy (as the resolute representative of a far-fetched and, for that very reason, unacceptable dualism) and—something which will be dealt with in due course—Heidegger's appraisal in cognitive science, which more often than not is understood as a mystical threat unable to contribute anything constructive to the field. That is to say, Wheeler wants to criticize the Cartesian assumptions underlying orthodox cognitive science but he will neither simply interpret Descartes's philosophy drawing heavily from Heidegger's own critique, nor interpret Heidegger, *à la* Dreyfus, as the staunch critic who would never accept the theoretical possibility of cognitive science. It must be possible to do both: to show Descartes's pervading influence on cognitive science and to embrace simultaneously Heideggerian insights without subscribing *tout court* to the consequences of his philosophy. This is precisely what Wheeler attempts to do.

Now, first a few words on Descartes. According to the standard interpretation of Descartes, the French philosopher was a substance dualist who ascribed preeminence to the inmaterial *res cogitans* over the *res extensa* and whose work has been easily superseded by the contemporary developments of cognitive science. If the mind is inmaterial, then it follows that it cannot be subjected to scientific inquiry. But Wheeler shows how this picture is rather a simplification of Descartes and why he should not be underestimated, since he was actually one of the early proponents of the mechanization of mind. As a matter of fact, quoting from Descartes's *Traité de l'homme*, Wheeler argues that "Descartes

takes a range of capacities that many theorists, even now, would be tempted to regard as psychological in character, and judges them to be explicable by appeal to nothing more fancy than the workings of the bodily machine" (2008, pp. 312-313). It is but stressing the accent on Descartes's dualism what has given rise to covert Cartesians within cognitive science, since it does not suffice simply to claim, for instance, that Descartes's neurophysiology was wrong—which it certainly was—nor even that it would be preposterous to suggest a similarity between the Cartesian soul and the contemporary scientific attempts at a theory of consciousness. The point is rather that deep explanatory principles and assumptions underlying work in contemporary cognitive science are decisively Cartesian. Therefore, one had better map out the course of these presuppositions of Cartesian descend.

On Wheeler's view, a series of Cartesian principles comprising Descartes's cognitive psychology must be identified in order to investigate as to what extent they remain parasitic in cognitive science. The Cartesian foundations bedeviling the cognitive enterprise are, however, exhibited by Wheeler as entering a period of quite dramatic reconstruction (for instance, in dynamical systems research), and his idea is that this reconstruction requires a more fundamental transformation in the philosophical foundations of the discipline. Wheeler's argumental strategy consists in pointing out eight principles of Cartesian psychology, and then argue that these principles "define a conceptual position toward which orthodox cognitive science tends overwhelmingly to gravitate, and at which it regularly comes to rest" (2005, p. 56). These principles include the following claims:

· The subject-object dichotomy is a primary characteristic of the cognizer's ordinary epistemic situation (p. 23).

· Mind, cognition, and intelligence are to be explained in terms of representational states and the ways in which such states are manipulated and transformed (p. 24).

· The bulk of intelligent human action is the outcome of general-purpose reasoning processes that work by (i) retrieving just those mental representations that are relevant to the present behavioral context, and then (ii) manipulating and transforming those representations in appropriate ways so as to determine what to do (p. 38).

· Human perception is essentially inferential in nature (p. 42).

· Perceptually guided intelligent action takes the form of a series of sense-represent-plan-move cycles (p. 43).

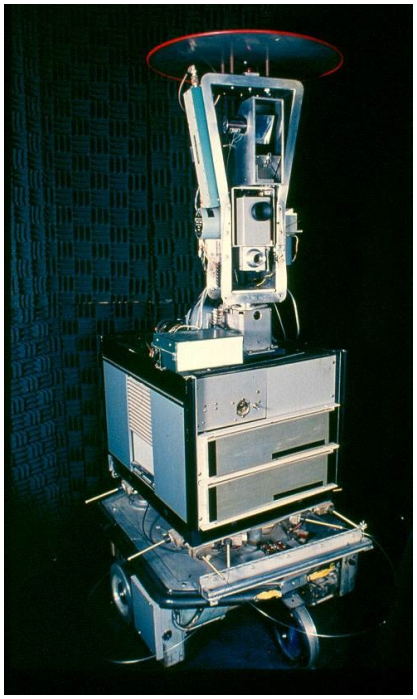
· In typical cases of perceptually guided intelligent action, the environment is no more than (i) a furnisher of problems for the agent to solve, (ii) a source of informational inputs to the mind (via sensing), and, most distinctively, (iii) a kind of stage on which sequences of preplanned actions (outputs of the faculty of reason) are simply executed (p. 45).

> · Although the informational contents carried by bodily sensations and certain primitive perceptual states may have to be specified in terms that appeal to particular bodily states or mechanisms, the cognitive-scientific understanding of the operating principles by which the agent's mind, given that information, then proceeds to generate reliable and flexible intelligent action remains conceptually and theoretically independent of the scientific understanding of the agent's physical embodiment (p. 51).
>
> · Psychological explanation is temporally austere, in that it is neither necessary nor characteristically typical for good scientific explanations of mental phenomena to appeal to richly temporal processes (p. 53).

It goes without saying that Wheeler is not arguing that the aforementioned principles of Cartesian psychology are shared *verbatim* in its entirety by every program and model in cognitive science. The 'Cartesian-ness claim'—the claim that orthodox cognitive science is founded on Cartesian principles—is that it is a modern species of Cartesian psychology. So as it happens, the neurophysiological facts between what we now know and what Descartes knew might differ, but the conceptual framework can, however, be kept intact. Accordingly, the Cartesian principles do not need to be shared identically but it must be possible to present evidence that each of the eight principles of Cartesian psychology are either "(i) an assumption made by at least the vast majority of orthodox cognitive scientists, ahead of the business of constructing specific explanations, or (ii) a core feature of certain paradigmatic, influential, or flagship examples of orthodox cognitive-scientific research" (Wheeler 2005, p. 56).

Typically, systems designed from the standpoint of orthodox cognitive science take perceptually guided intelligent action to be a series of sense-represent-plan-move cycles (Wheeler 2005, p. 67). The architectural blueprint behind this conception of action is what Brooks called the sense-model-plan-act (SMPA) framework (1999a): a cognitive architectural framework that Brooks rejects for it dissociates perception and action. Take Shakey: a robot designed in the late 1960s at the Stanford Research Institute (Nilsson 1984) that embodies the typical technological implementation committed to SMPA. Shakey has visual images that are received via a black-and-white television monitor, it represents images based on a model of the world built as a set of first-order predicate calculus, and the world model is delivered to a central planning system called STRIPS based on a GPS architecture. Finally, in order to move, STRIPS-generated action-specifying expressions are decoded into a format appropriate for driving the motors (Wheeler 2005, p. 69). For generating action, Shakey must appeal to its internal world-model, where conventions have been established for representing doors, wall faces, rooms,

objects, and the robot's status (Nilsson 1984, p. 19). Under such robotic world-modeling, the environment is understood as no more than a source of problems, obstacles, ambience information, and settings for action, but not as constituting cognition or playing any distinctive role in the formation of intelligence. From this follows an important conclusion for the philosophy of mind pervading orthodox cognitive science: what is important is to be located in the head. Hence cognition is essentially intracranial.



Shakey, manufactured by SRI International.

A further example provided by Wheeler as evidence of the 'Cartesian-ness' of traditional cognitive science is Marr's computational investigation into the human representation and processing of visual information (2010), according to which the underlying task of vision is to reliably derive properties of the world from images of it. On Poggio's view, "the central tenet of Marr's approach is that vision is primarily a complex information processing task, with the goal of capturing and representing the various aspects of the world that are of use to us" (1981, p. 3). Most importantly, for Marr these representations are to be conceived of as context independent, given that they are not the product of any fundamental embedding or a result of the specific context of action, but rather of mere neural activity. This intracranial environmentally detached 'neurocentrism' is an essential tendency at work within orthodox cognitive science:

> According to such a view, although both the agent's nonneural body (e.g., muscular adaptations, the geometric properties of limbs) and the agent's environment are clearly essential, in some sense, for intelligent action to occur as it does, the processes that account for the richness and flexibility that are distinctive of such behavior remain fundamentally neural (e.g., neurally realized mechanisms of inference, discrimination, estimation, and route planning). Put another way, the message is that the causal factors that explain the adaptive richness and flexibility of naturally occurring intelligent behavior are located neither in the agent's nonneural body nor in her environment, but pretty much exclusively in her brain. (Wheeler 2005, p. 81)

The dispute behind this rejection of neurocentrism is between contingent intracranialism and contingent transcranialism (Adams & Aizawa 2009). According to the former, tool use—a commonplace in situated cognition and the extended mind hypothesis[97]—is a matter of intracranially localized cognitive processes interacting with extracranial biological, chemical, and physical processes (Adams & Aizawa 2009, p. 78). So vision, for example, is essentially cognitive and begins in the retina. The transcranial approach, on the other hand, emphasizes noncognitive processes which span the cranial boundaries and extend into extracranial space. As such, due to its noncognitive emphasis, it tends to be regarded by intracranialists as a threat to cognitive science. Whereas for the former approach meaning is decisively in the head, for the latter "we do not store the meaningful inside ourselves, but rather live and are at home in it" (Haugeland 1998, p. 231). The intracranialist would reply that that is beautiful poetry, because the condition for living and being at home in such a way presupposes the primary contribution of neural activity.

It is true that Descartes did not (and could not) have any developed neuroscientific knowledge at his disposal. Rather, his explanations were modeled on the basis of a system of hydraulics: nerve fibers stretching organs, tensions and relaxations closing or opening brain cavities releasing a flow of animal spirits through a corresponding point of the pineal gland (Wheeler 2008, p. 310). But that does not annul the fact that, on Wheeler's words, "if we shift to a more abstract structural level of description, what emerges from that theory is a high-level specification for a control architecture, one that might be realized just as easily by a system of electrical and biochemical transmissions—that is, by a system of the sort recognized by contemporary neuroscience—as it is by Descartes's ingenious system of hydraulics" (2008, pp. 310-311). The point is that neurocentrism is committed to the

[97] The fundamental reference regarding the extended mind hypothesis is, of course, the famous paper penned by Clark and Chalmers titled precisely 'The Extended Mind' (2010; originally published in 1998). This hypothesis is the first formulation of transcranialism: "an *active externalism*, based on the active role of the environment in driving cognitive processes" (Clark & Chalmers 2010, p. 27). In another paper, Adams and Aizawa (2010) have defended the bounds of cognition, that is, the fundamental intracranial character of cognition.

Cartesian view of the explanatory disembeddedness of intelligent action, whereby the environment is one of two things, or both at the same time: (i) a furnisher of problems for the agent to solve (as exemplified by the typical tasks entrusted to robots designed on the basis of the SMPA framework, like Shakey), or (ii) a source of informational inputs to the mind via sensing (as exemplified in Marr's computational theory of vision). From the standpoint of orthodox cognitive science, explaining vision is understanding how the brain builds up an internal model of the world based on light reception and inference from external traits. In contrast, "the subject of the new vision science is explaining why it seems to us as if the brain does this, when in fact it does not" (Noë 2002, p. 140). Consequently, Noë —another apostle of the new science of mind—defends a radical anti-Cartesianism: "my position is simple: Cartesian neuroscience has no empirical support for its basic assumption that conscious experience is an exhaustively neural phenomenon" (2002, p. 173).

So Descartes stands firm in the middle of this dispute. He can certainly be mocked for postulating the pineal gland as the interface between mind and matter, where physical stimuli are converted into perceptions and representations, and where motor instructions are turned into physical processes due to its involvement in important processes such as imagination, sensation, memory, and the causation of bodily movements. This is the typical arrogance of our contemporaries: they can look back in the past and laugh at how wrong philosophers of foregone epochs got the facts. However, Wheeler reminds us that even though for contemporary orthodox cognitive scientists Descartes's pineal gland does not exist and even arouses derision, an explanatory interface responsible for the conversion of physical stimuli into representational and perceptive processes does! (2005, p. 85). That is to say, the refinement of the facts does not revoke the underlying explanatory framework. On Wheeler's account, today's pineal gland can be located over the point of the body where sensory transducers convert physical stimuli into representational states and where motor transducers convert representational states into physical processes producing bodily motions (*idem*). To say it with Haugeland, transducers are AI's answer to the pineal gland: "the relevant transductions would have to take place *within* the brain, between one part of it and another—not so far from the pineal gland, as luck might have it (1998, p. 228). Central to this underlying explanatory framework is the Cartesian idea that the mental is an ontologically independent domain—and hence the need for transducers doing the coupling between

the physical and the mental. But it must be granted that this necessity is an artefact arising from the assumed Cartesian framework. It seems that there has to be a conversion between the symbolic contents of the mind and the physical processes of the body, but only because the separation between the ontological domains is supposed to be fundamental. The assumption of neurocentrism as well forces one to conceive of an intelligent agent as responding primarily to internal representations when coping, rather than to the environment or to the world, which most certainly biases the orientation toward representation and away from perception.

Looked more closely, not only abstracting Dasein from the worldhood of the world (*Weltlichkeit der Welt*) is preposterous. Take, for instance, the move or play in a game as pointed out by Haugeland:

> pushing around a little piece of plastic shaped like a turret could only amount to a *rook* move in an appropriate spatial and temporal context of other chess pieces and moves. To call it a rook move apart from such context is simply nonsense. Likewise, so the reasoning goes, to regard any phenomenon as intentional or normative in isolation from the relevant whole, is also nonsense. And since, in the case of mental attribution, the relevant whole must include the individual's environment and/or community, the Cartesian independence of the mental is impossible. (1998, p. 208)

It should then not come as a surprise that Heidegger's phenomenology of world is invoked in conjunction with the attempt to escape from the Cartesian mindset. The idea of a neo-Heideggerian framework, from whence to rethink the cognitive enterprise away from Cartesian ways of thinking, has cavorted Wheeler's mind at least ten years (see 1995) prior to publication of his *Reconstructing the Cognitive World* (2005). But Wheeler's idea is not only critical but, above all, constructive. His framework is neo-Heideggerian because Wheeler cannot simply buy the 'anthropocentric tendencies' in Heidegger's philosophy (1995, p. 69) and wants to extend accordingly the consequences of the Heideggerian framework to the animal kingdom. This is due to the *Muggle constraint*, which guides obligingly Wheeler's investigation:

> In J. K Rowlings's Harry Potter books, there are two coexisting and intersecting worlds. The first is the magical realm, populated by wizards, witches, dragons, dementors, and the like. This is a realm in which, for example, getting from A to B can be achieved by flying broomstick, flying carpet, or more dramatically, teleportation, and in which one object can be transformed into another by a transfiguration spell. The second world is the nonmagical realm, populated by Muggles—Muggles like us. Muggles, being nonmagical folk, are condemned to travel by boringly familiar (to us) planes, trains, and automobiles, and to operate without the manifest benefits of supernatural object-altering powers. Now, if you want to understand of how Muggles work, you had better not appeal to anything magical. (2005, p. 4-5)

The point is simple: "no spooky stuff allowed" (2005, p. 5). So Wheeler is ready to incorporate Heidegger's phenomenology of agency within the context of the environing world (*Umwelt*) to bear on

cognitive science, but he is not willing to abandon what he calls an intellectual marriage of philosophy and science (2005, p. 4). No 'spooky stuff' means also, in Wheeler's terms, that no occult entities unbeknownst to scientific inquiry are to be let in: "if philosophy and natural science clash (in the sense that philosophy demands the presence of some entity, state, or process that is judged to be inconsistent with natural science), then it is philosophy and not science that must give up" (2005, p. 5). This can be couched in the following terms: philosophy is all right, but in this intellectual marriage demanded by Wheeler, it is science the one with preeminence when it comes to the postulation of existing objects. Escaping the Cartesian mindset does not impose on us an obligation of forsaking the grounds of a scientifically informed philosophy. We are, after all, Muggles, and must abide to nonmagical methods and to the constraints they impose.

For the aforementioned intellectual marriage between philosophy and science to be effective, the phenomenology of agency in the environing world cannot simply have emerged out of nothing. On Wheeler's account, to suppose so would be 'spooky.' So somehow sense has to be made about how such an instance of meaningful structures, like Dasein's environing world, has arisen from more primordial biological structures. According to Wheeler, evolutionary biology provides a treatment of other animals which does not consider them under the light of value-free objects (that is, as present-at-hand entities) but rather as creatures immersed in their natural niches that can also be said to carry out intelligent, at least meaningful, activities. On Wheeler's view, Heidegger's conception of animals is not far from Descartes's derogatory idea that animals are mere automata. It is no secret that, for Heidegger, animals have no world and do not exist in the sense of Dasein's existence. Instead, animals are 'world poor' (*weltarm*) and are absorbed in captivation (*Benommenheit*):' "captivation is the condition of possibility for the fact that, in accordance with its essence, the animal *behaves* [benimmt sich] *within an environment* [Umgebung] *but never within a world* [Welt]" (GA 29/30, pp. 347-348). But this Heideggerian opposition between *Umgebung* and *Welt*—which Heidegger imports from biologist Jakob von Uexküll—appears rather extravagant and untenable to Wheeler, who does not shy away from accusing Heidegger of anthrochauvinism (2005, p. 157). Perhaps the human *Welt* is more elaborate than the animal *Umgebung*, but one cannot be dismissive of the fact that humans are also animals.

This is the step forward towards naturalism that Wheeler invites his readers to take: the world has to be somehow a continuation of the biological environment. An abismal gap between Dasein and other animals can be granted on the basis of human cultural evolution, but it cannot be forever unsurmountable, for Dasein was once a 'mere' animal. Dasein, to be clear, is still an animal, albeit provided with a set of tools for thinking which function as imagination-extenders and focus-holders (Dennett 2013). Neither can be supposed that all living beings are encapsulated in their own irreductible worlds which do not communicate with other worlds—as Jakob von Uexküll (2010) would have it—since evolution requires continuity between species, interrelationship, and kinship between all species; an idea that von Uexküll would nevertheless pin on the illusion founded on the belief in an unitary world shared by all existing creatures (Agamben 2002, p. 50).[98]

Rejecting Heidegger's claim to human exceptionalism, Wheeler now turns to a difference between the physical and the biological sciences that can be couched in terms of a distinctive treatment toward behavioral ecology: "this area of biology (at least) cannot be treated as equivalent to the physical sciences. In fact, it seems that for the discipline of behavioral ecology to make sense, the capacity of animals to open up domains of significance has to be assumed" (Wheeler 1995, p. 72). And again, even if these domains are not as rich and manifold as the ones opened up by humans, they are nonetheless evolutionary antecedents of human existence. Wheeler's argument is the following: one can certainly concede that there is an abstraction of meaningful contexts in physics, for which *Vorhandenheit* can, no doubt, be said even to be its sole business, but this is incorrect in the case of living organisms. After all, it is thanks to evolutionary theory that a continuity between animal species must be granted, so that human intelligence—for all its distinctiveness—can be traced back to more primitive forms of interaction with the environment. Human beings and their creations (including knowledge, language, and morality) are a byproduct of evolutionary history.[99] The meaningful structures constituting Dasein's world are no exception. Wheeler has it that even if we consider living organisms to be objects, our scientific findings are not up to us, since animals are "autonomous agents who adopt strategies with

[98] On von Uexküll's view, animal worlds are not only unknown worlds but, more importantly, they are invisible and inaccesible from our own human perspective. A mechanistic view of nature contributes to this inaccessibility: "whoever wants to hold on to the conviction that all living things are only machines should abandon all hope of glimpsing their environments" (2010, p. 41).

[99] For a succinct account of evolutionary theory bearing on knowledge, language, and morality, see Ruse (2012).

fitness consequences for both the strategy adopting animal itself, and the other animals with whom it interacts" (*idem*). We might want to theorize animals as if they were mere objects, but they are rather living organisms with coping and survival strategies. They might be *weltarm*, but they are not *weltlos*. For this very reason, we cannot conceive of animals as though we were dealing with inanimate things, for our objective (present-at-hand) stance towards them is just our stance, but not theirs. A real science of living organisms must provide a detailed account of how animals behave and develop, and not just envisage them as something they are not. So this explains Wheeler's piecemeal approach to Heideggerian philosophy: Dasein, as an evolved creature, has also had to learn to cope with its world and this is why it would be a mistake to attempt to abstract it away from an evolutionary approach. Now it is clear why Wheeler's approach is neo-Heideggerian: it accepts Heidegger's phenomenology of Dasein's world as an accurate description of the structures of coping within the practical context of human agency, but Wheeler does not see why simpler forms of animal intelligence cannot be integrated into a more encompassing picture. What impedes one from conceiving of meaning as pervading the worlds of other animals? Why would anyone think that Dasein is the only exception in the whole of the animal kingdom?

It is worth remarking that this line of thinking can also be applied to artificial life (A-Life for short) research: the field of study associated with systems related to life, its processes and evolution by means of computer simulations, robotics, and biochemistry. Wheeler takes A-Life in the restricted sense of being an attempt to explain evolutionary and adaptive systems, including phenomena customarily grouped together with labels such as intelligence, mind, and cognition (1995, p. 65, n. 1). Wheeler addresses this possibility via the example of cellular automata. According to Varela, Thompson and Rosch (1993), Bittorio, a ring of eighty elementary processing units (cellular automata) in a random soup of 1s and 0s endowed with a network architecture and coupling properties, can illustrate how very simple organisms—even artificial ones—can begin to enact a world by exhibiting emergent properties. As a cellular automata, Bittorio can demonstrate various kinds of fantastic emergent behavior according to the state of neighboring cells, since the state of one of its cells is replaced by a perturbation when one of the two alternatives (0 or 1) is encountered. So the experimenter needs only to stipulate the possible states into which each unit is able to move, the rules

governing the way the units change as a result of local interactions with neighboring units, and the way in which the network is coupled to a random milieu. What is interesting is that a Bittorio "picks up or singles out from the milieu a very specific subset, namely, finite odd sequences, since only these sequences induce a repeatable change in Bittorio's configuration" (Varela, Thompson & Rosch 1993, p. 152). The conclusion is rather surprising: "given its rule and given its form of structural coupling, this Bittorio becomes an odd sequence recognizer" (*idem*). For what is worth, Wheeler thinks that here the system's activity can be understood as a sort of minimal interpretation, meaning that Bittorio "selects or brings forth a domain of significance out of the background of its random milieu" (Varela, Thompson & Rosch 1993, p. 156). The authors do not refrain a single step back from speaking of 'Bittorio's world:'

> It should be apparent that this world is not pregiven and then recovered through a representation. We did not design Bittorio to be an odd sequence recognizer; we simply provided Bittorio with certain internal dynamics and then dropped it into a random milieu. Nevertheless, given the history of coupling between the internal dynamics and the milieu, *odd sequence* becomes a significant distinction for Bittorio. For this reason, we describe Bittorio's world as enacted through a history of structural coupling. (Varela, Thompson & Rosch 1993, p. 156)

According to Varela, Thompson and Rosch, Bittorio can be conceived of as "a minimal example of how an autonomous system brings forth significance from a background" (*idem*). The point reinforces Wheeler's idea that such a simple system's activity performs a minimal kind of interpretation and, what is more interesting, it might help us understand how an agent's ongoing activity brings forth significance from an essentially meaningless background (Wheeler 1995, p. 70).

So basically, the neo-Heideggerian approach demands that Heideggerian philosophy meet the Muggle constraint. Orthodox cognitive science restricted itself to offline intelligence, that is, to phenomena involving propositional knowing-what, like weighing up the pros and cons of carrying out a certain action in a specific situation (Wheeler 2005, p. 12). In contrast, the new approach is ready to do both: subsume GOFAI by allowing it an authoritative place in offline phenomena research (Wheeler 2005, p. 249), while at the same time investigating the largely unexplored dimension of online intelligence comprising the sort of phenomena characteristic of agents that produce a suite of fluid and flexible real-time adaptive responses to incoming sensory stimuli. And this sort of intelligence—albeit more complex and sophisticated in humans—is not a privilege of Dasein alone but is scattered in the chain of being, from nonhuman animals to cellular automata.

The present chapter and the previous one have reviewed in depth the main ideas surrounding the appropriation of Heidegger within contemporary cognitive science. It can scarcely be said that these approaches are objectionable from a purely Heideggerian point of view. Therefore, the next section will attempt to exhibit the kinds of objections that can be presented to the 'Heideggerian alternative' in cognitive science. But this shall not only be attempted from the viewpoint of 'pure' philosophical musings. The point is rather to show that the naturalization of Heidegger—and the naturalization of intentionality which underlies it—is a glaring error which belittles human experience. This results from an analytic reception of Heidegger. 'Analytic' not only in the sense that Heideggerian philosophy is appropriated by analytic-trained Anglo-American philosophers, but also in the decisive sense that the Heideggerian philosophy which is appropriated for the purposes of advancing the new paradigm, pays only attention to specific parts of Division I of *Sein und Zeit*; parts which, in the same vein, are also appropriated very selectively. The reception is 'analytic' in that it constitutes a very schematic version of Heidegger taking precisely his thought out of context (Rehberg 2012, p. 160). In order to complete the purposes of this chapter, I now turn to the aporias of what can be deemed the 'analytic' reception of Heidegger's philosophy. This will let us exhibit some criticisms directed not only to Wheeler's approach explained in this chapter, but crucially towards the whole spectrum of the so-called 'Heideggerian alternative,' which includes the approaches reviewed in the previous chapter.

Aporias of the 'Analytic' Heidegger

The heavy import of Heideggerian vocabulary into the new movement in cognitive science has, of course, not gone unnoticed for Dreyfus. Dreyfus's position on the so-called 'Heideggerian alternative' is, however, rather critical. According to Dreyfus, the different approaches presented above are plagued with problems and misunderstandings, although he is enthused over a "positive account of how Heideggerian AI and an underlying Heideggerian neuroscience could solve the frame problem" (2007a, p. 254). Accordingly, Dreyfus criticizes Brooks because "his robots respond only to *fixed features* of the environment, not to context or changing significance. They are like ants, and Brooks aptly calls them 'animats'" (2007a, p. 250). So Brooks's idea that his creatures respond to the world itself is somewhat exaggerated.

More objectionable is perhaps that Brooks renounced to his initial modesty according to which it was reasonable first to focus on insect-like animats to later on advance in the design of more complex robots. The Cog project was such leap ahead from simple insect-like devices to a full-blown humanoid, with the inherent theoretical difficulties which turned subsequently into unsolved problems, and that led to Cog being exhibited in a museum as what it is: past history of a fanciful project. Dreyfus despairs of Dennett (the co-worker with Brooks on the Cog project) for not having recognized openly that something went completely wrong in the project, let alone that it was absurd given the knowledge available. Instead, Dennett merely noted in a personal communication to Dreyfus that "progress was being made on all the goals but slower than had been anticipated" (Dreyfus 2007a, p. 250). For Dennett, the problem ultimately boils down to lack of sponsors and money. This is also an opinion expressed by Minsky in several interviews: when it comes to AI, the prospects are theoretically difficult and, for that very reason, understandably slow, so funding for research in artificial architectures begins to grow scarce when results are to be expected realistically decades, even centuries ahead, which is only natural. In addition, AI "ran out of high-level thinkers" (Minsky 2007b), thinkers, that is, willing to embark on the decisive journey of using the sciences of the artificial to understand natural intelligence —not just concerned with projects suited for industrial purposes. Brooks, however, is more willing to accept that "perhaps at this point we simply do not *get it*… there is some fundamental change necessary in our thinking in order that we might build artificial systems that have the levels of intelligence, emotional interactions, long term stability and autonomy, and general robustness that we might expect of biological systems" (1997, p. 301). For Dennett and Brooks as well, the issue is simply this: human intelligence is a continuum with insect intelligence, therefore, as Dreyfus notes, "adding a bit of complexity to what has already been done with animats counts as progress toward humanoid intelligence" (2007a, p. 250). Dreyfus seems to think that there is nothing in the least reprehensible in the research program involving animats or even cellular automata, since such research might shed some light on how basic intelligence works from the bottom-up. However, one must guard against committing the first-step fallacy (Dreyfus 2012) and succumbing to the view that one can so easily bridge the complexity gap between creatures that are not even alive and human beings. If Heidegger conceives of an abyss (*Abgrund*) separating Dasein—who is *weltbildend*—from the animal—which is

weltarm—the gap is more pronounced and even unsurmountable with unliving creatures—which must certainly be *weltlos* (GA 29/30). But still, it is interesting to note that Dreyfus also thinks this is *just* a gap involving complexity. The problem with projects such as Cog is the absurd optimism under which such endeavors are premised, not—contrary to that with which Dreyfus has always been charged of as defending—their purported unrealizability. A human being is simply a conglomerate of many complicated systems including a bodily wetware capable of skillful learning and performance. We are nowhere near the possibility of computationally designing a program capable of skillful coping in such complexities like the ones presented by the human world. But, again, there is no in-principle argument against the possibility of discovering the underlying mechanisms giving rise to such meaning-laden spectacle.

Agre's ideas seem to be more appealing to Dreyfus than Brooks's, and so Dreyfus even credits Agre with understanding *Zuhandenheit* better than he himself did, since for Agre ready-to-hand is not a *what* but a *for-what* (2007a, p. 252). So Dreyfus has it that Agre was able to show how Heidegger wants to get at something more basic than simply a class of objects: equipment (*Zeug*). The entire point of the equipmental character of things in the world is not that they are entities with a function feature—this was Dreyfus's previous interpretation of *Zeug* and *Zeugzusammenhang*—but rather that they open up possibilities for action, solicitations to act, and motivations for coping; an idea that Dreyfus takes admittedly from Agre's endeavors towards modeling *Zuhandenheit* on the basis of deictic intentionality. Nevertheless, Dreyfus is of the opinion that in attempting to program ready-to-hand, Agre succumbs to an abstract objectification of human practice, because affordances—inasmuch as they are not objects but the in-between interaction in which no subject nor object is involved—are not amenable to programming. That they are not is not something Agre seems to fully understand, and this is why he thinks that somehow deictic representations must be involved in human understanding. According to Dreyfus, "Agre's Heideggerian AI did not try to program this experiential aspect of being drawn in by an affordance. Rather, with his deictic representations, Agre *objectified* both the functions and their situational relevance for the agent. In *Pengi*, when a virtual ice cube defined by its function is close to the virtual player, a rule dictates the response (e.g., kick it). No skill is involved and no learning takes place" (2007a, p. 253). It must be admitted that a virtual world is not even slightly comparable with the

complex dynamics of the real word. In a virtual world, the dynamics of relevance are determined beforehand, so a program like *Pengi* simply cannot account for the way human beings cope with new relevancies. Like Brooks, concludes Dreyfus, Agre "finesses rather than solves the frame problem. Thus, sadly, his Heideggerian AI turned out to be a dead end. Happily, however, Agre never claimed he was making progress towards building a human being" (2007a, p. 253).

After leaving behind the 1980s turn to embodiment and the 1990s attempt at putting to work for the first time Heideggerian insights in computer programming, Dreyfus turns towards the most recent appropriation of Heideggerian philosophy within contemporary cognitive science research, namely Wheeler's approach, which he christens rather critically 'pseudo Heideggerian AI.' According to Dreyfus's assessment, this approach has a fake character which is revealed by its use of representations. And this sole fact makes it hard for one to deem it really Heideggerian in scope. Indeed, Wheeler presents his global project as requiring "a defense of action-oriented representation… action-oriented representation may be interpreted as the subagential reflection of online practical problem solving, as conceived by the Heideggerian phenomenologist. Embodied-embedded cognitive science is implicitly a Heideggerian venture" (2005, p. 222-223). Wheeler admits to having been influenced by Dreyfus's critique of artificial reason, but not without adding one important caveat: Dreyfus's focus on the problem is wrong because his position is a controversial negative assessment of the empirical achievements of orthodox AI peppered with arguments against Cartesianism. So more than wrong, Dreyfus's critique of artificial reason is incomplete. That Dreyfus's critique is inconclusive can be confirmed in that it does not offer a way out from the most recalcitrant theoretical shortcomings of the traditional approach. So Wheeler has it that Dreyfus's is a purely philosophical take on AI, since a truly cognitive approach should offer solutions for those aspects of the traditional way of doing things. If a mistake is shown—so seems to be Wheeler's argument—then solutions leading to improvement must be presented as well. Therefore, "it is not any alleged empirical failure on the part of orthodox cognitive science, but rather the concrete empirical success of a cognitive science with Heideggerian credentials, that, if sustained and deepened, would ultimately vindicate a Heideggerian position in cognitive theory" (2005, p. 188-189).

However, Dreyfus doubts of the Heideggerian credentials of Wheeler's approach and retorts that "merely in supposing that Heidegger is concerned with subagential *problem solving* and action oriented *representations*, Wheeler's project reflects not a step beyond Agre but a regression to pre-Brooks GOFAI" (2007a, p. 254). Dreyfus's critique of Wheeler's use of Heideggerian philosophy consists in pointing out that being-in-the-world is more basic than thinking and solving problems. Wheeler's preoccupation with how to accommodate Heideggerian insights into a representational framework is just an artefact created by his theoretical convictions, which are cognitive all too cognitive in that they suppose as fundamental the existence of a mind which essentially characterizes what human beings are. Wheeler's anthropological motto seems to be thus: *human being is cognitive, therefore his essence lies in cognition*. In contrast, cognition is for Dreyfus just a derivative product owing its existence to the more fundamental social dealings of a practical coper: the human agent concerned with her own existence.

However, it is important at this juncture to draw more general conclusions from the appropriation of Heidegger's philosophy in cognitive science. Brooks, Agre, Wheeler, and Dreyfus all share at least the idea that something like a Heideggerian cognitive science and a Heideggerian neuroscience is possible (although Brooks would probably not want anything to do with philosophy). That Heidegger's philosophy or phenomenology in general can enter in dialogue with cognitive science implies a certain capitulation to some sort of naturalism. In the work of neuroscientist Walter Freeman (1999), for instance, Dreyfus finds the key for liberating the mind—or in this case, the brain—from the necessity of representations. As such, this is a true step forward, according to Dreyfus, in the right direction of founding a Heideggerian neuroscience. Let us recall that the brain, according to Freeman, is a nonlinear dynamical system which can find and augment significance in the world. On Dreyfus's interpretation of Freeman's neurodynamics,

> the important point is that Freeman offers a model of learning which is not an associationist model according to which, as one learns, one adds more and more fixed connections, nor a cognitivist model based on offline representations of objective facts about the world that enable offline inferences as to which facts to expect next, and what they mean. Rather, Freeman's model instantiates a genuine intentional arc according to which there are no linear causal connections nor a fixed library of data, but where, each time a new significance is encountered, the whole perceptual world of the animal changes so that the significance as directly displayed is contextual, global, and continually enriched. (2007a, pp. 260-261)

His great expectations concerning such project in neuroscience explains why in the end Dreyfus has not been able to resist a sort of naturalism according to which being-in-the-world can be traced back to

perception-action loops taking place in neural activity. Dreyfus might doubt of the Heideggerian credentials of other philosophers, but his are in no way established. For Dreyfus, the point is to get the right underlying neuroscience with Heideggerian credentials and thus he seems to ignore that such expectations are not phenomenological. This agreeable approach with Heideggerian credentials, Dreyfus's finds in Freeman's neurodynamics, since Freeman conceives of intentionality as requiring "acting to create meaning instead of just thinking" (1999, pp. 38-39). On Freeman's view, neuronal processes are not just computational tasks being carried out by billions of semantically blind nanobots, but are instead in themselves meaningful-creating processes. This idea—which Freeman thinks is shared by philosophers like Heidegger and Merleau-Ponty, by psychologists like Gibson, and pragmatists like Dewey—"is crucial for interpreting my observations in experiments on neural mechanisms of perception in animals" (1999, p. 40). Heidegger's appropriation in neuroscience consists in supposing that the structures comprising Dasein's world can somehow be brought to bear on systems that are found in the natural world. This is, incidentally, a perfect contradiction to Heidegger's Uexküllian conviction that the world is exclusively a human phenomenon.[100]

The preceding points bring to the fore the problematic character of Dreyfus's hopes for a Heideggerian neuroscience. He seems to think that the difference between machines, animals, and humans can be sorted out, if not by sheer adding a bit of complexity to computational machinery, at least by parting from the right neuroscience and then from a correct conception of embodiment and embedding: "to program Heideggerian AI, we would not only need a model of the brain functioning underlying coupled coping as Freeman's, but we would also need—and here's the rub—a model of our particular way of being embedded and embodied such that what we experience is significant for us in the particular way that it is" (Dreyfus 2007a, p. 265). So Dreyfus sets a high standard for the Heideggerian alternative to be fruitful, but the possibility of designing a program of being-in-the-world is not declared as forever unviable, nor is the underlying assumption that being-in-the-world starts in very basic natural processes, ultimate denizens of brain activity, denied intelligibility. The typical

[100] To be more precise, Jakob von Uexküll holds that "every animal, no matter how free in its movement, is bound to a certain dwelling-world…" (2010, p. 139). This dwelling-world is incommensurable with the human world. Therefore, Heidegger's denial of a shared world between humans and other animals reverberates with Uexküllian overtones.

objection against Dreyfus that his critique of artificial reason is just an in-principle argument is then completely false.

However, there are more critical factors at work from the phenomenological point of view than Dreyfus has been able to reckon with. As a matter of fact, according to a fundamental phenomenological insight, there is something profoundly misleading in attempting to naturalize consciousness. Indeed, Husserl has been vocal over and over again against the perverseness (*Verkehrtheit*) of the naturalization of consciousness, because it obfuscates (*macht blind*) "not only the I but everything that is characteristic of consciousness" (Hua VII, p. 105 ff). The subjective is, for Husserl, unreal from the point of view of the real ("gegenüber der Realität eine Irrealität," [Hua IV, p. 64]). Moreover, "as long as naturalism suffices, there reigns… the theoretical blindness for the specificities of the mind [*das Spezifische des Geistes*]" (Hua XXXVII, p. 122 ff). In sum, Husserl does not dither from speaking of a nonsensical naturalization of the mind (*widersinnigen Naturalisierug des Geistes*) or a reification of consciousness (*Verdinglichung von Bewusstsein*). Both naturalism in the form of the natural sciences and historicism and worldview philosophy (*Weltanschauungsphilosophie*) naturalize respectively consciousness and the ideas. As such, naturalism and historicism are nonscientific: not *per se* but only in reference to their absolutism regarding their own methods (Rinofner-Kreidl 2000, p. 752).

Be that as it may, Dreyfus conceives (wrongly) of a profound breakthrough between his beloved Heidegger-style phenomenology and Husserl's, and so goes on to affirm that since Husserl "put directedness of mental representations at the center of his philosophy, he is also beginning to emerge as the father of current research in cognitive psychology and artificial intelligence" (1982, p. 2). But given the aforementioned Husserlian critique of the naturalization of mind and consciousness, it strikes one as puzzling how Husserl can seriously be considered the father of cognitive psychology and AI. So if Dreyfus designates Wheeler's approach pseudo-Heideggerian, it must be granted that his own approach is not less *sui generis*, when not equally pseudo-phenomenological.

For similar reasons, it is difficult to see how Heidegger's phenomenology can be accommodated into a naturalistic framework that can meet Wheeler's Muggle constraint. It is Wheeler's idea, however, that the Heideggerian credentials of his project are fully intact (2005, p. 190) and that Heidegger's philosophy can serve as a 'conceptual adhesive' binding together a well-anchored research program

(2005, p. 191). In order to ground out this marriage between Heidegger's philosophy and cognitive science, Wheeler entertains a rather strange interpretation of *Vorhandenheit*. Although on occasion Heidegger himself seems to refer to present-at-hand in derogatory terms ("*sondern setzt… die Seinsheit purer Vorhandenheit nur wieder voraus*" [SZ § 21, p. 99], etc.), Wheeler does not think *Vorhandenheit* is to be interpreted with contempt. Instead, Wheeler has it that Heidegger has shown that scientific objects are discovered in the world as 'deworlded' entities, and because they are only the business of science alone, philosophy cannot claim a right to postulate occult entities unknown to science. If the business of science is the investigation of deworlded entities, that is, inasmuch as they are not dependent on human context bias, philosophy on its part must dedicate its efforts to world contextuality and thus can serve as an adhesive accompanying scientific endeavors. To put it bluntly: philosophy can clarify the world of the deworlded entities of empirical science. Hence, here lies a difference between philosophy and science which, according to Wheeler, must not be missed: a difference that is also an interplay between the two that clarifies as well how they relate to each other. Taking into account a distinction introduced by McDowell (1994), Wheeler maintains that empirical science provides an *enabling understanding*, which reveals the causal elements and the organization of the systematic causal interactions between those elements. In contrast, philosophy is characterized by a *constructive understanding*, which concerns the identification, articulation, and clarification of the conditions that determine what it is for a phenomenon to be the phenomenon that it is (2012, pp. 182-183).

Under the last distinction, Wheeler thinks that Heidegger's philosophy can be depicted as seeking a particular sort of constitutive understanding. Namely: "the understanding in question concerns an account of the conditions that determine what it means to live a human life" (Wheeler 2012, p. 183). And the recognition of this transcendental role of philosophy and the clarifications it brings forth might be also illuminating to empirical science. On Wheeler's view, constructive understanding does not have to be reduced to enabling understanding, for that would be tantamount to a kind of *scientism* (2012, p. 185). Rather, "our two kinds of understanding (and thus philosophy and cognitive science) will standardly engage in a process of mutual constraint and influence that McDowell tags with the enticing phrase a perfectly intelligible interplay" (*idem*). This view of the interplay between empirical science and philosophy and thus of the complementariness of phenomenology and cognitive

science is supplemented by a minimal naturalism: "because it stops a long way short of reductionism, minimal naturalism does not demand that a complete cognitive science of Dasein would be a complete understanding of Dasein, although it would be a complete enabling understanding" (Wheeler 2012, p. 191). At the behest of this interplay between the transcendental conditions presupposed by any empirical science and Wheeler's minimal kind of naturalism, it is possible to hold that transcendental conditions "are not immune to revision, or even perhaps rejection, in the light of the results of the empirical scientific research that they make possible" (Wheeler 2012, p. 192).

However, for the phenomenologist, this empirical revision of the transcendental sphere—the sole suggestion that empirical results might modify the transcendental conditions pervading our understanding—is as absurd (*das ist ein reiner Widersinn*, as Husserl would surely hold) as if "one demanded causal properties and relationships etc., for mathematical numbers" (PsW, p. 35). It is important to note that this Husserlian transcendental stance has also been thoroughly assumed by Heidegger in *Sein und Zeit*. Heidegger does not refrain from criticizing his phenomenological master from time to time, nor from secluding himself from the way Husserl practices phenomenology. But Heidegger is surely no pre-Husserlian philosopher, which most certainly means that his entire work is not even understandable without the impulses he received from Husserlian phenomenology. Precisely, Ratcliffe has criticized Wheeler's naturalization of Dasein by extracting the following transcendental argument from Division I of Heidegger's *magnum opus*:

> · Empirical science enquires as to what is the case. In order for it to do so, the distinction between being and not being the case must already be intelligible. In other words, one must have an understanding of what it is to *be*. This presupposed understanding need not be restricted to the Being of the present-at-hand, to the kind of Being that empirical science is concerned with, but it does at least include it.
>
> · We are the beings that have an understanding of Being… Hence any comprehensive account of human understanding needs to include an account of our understanding of Being.
>
> · Empirical science is concerned with the present-at-hand, rather than with what is presupposed by the intelligibility of presence-at-hand; it addresses only what kinds of things the world is populated with. Therefore it cannot incorporate an adequate account of Dasein.
>
> · In addition, empirical scientific theories cannot adequately encapsulate the having of a world, a characteristic that is inseparable of Dasein. The world we find ourselves in, which is made intelligible by our understanding of Being, is a world *in which* we encounter the present-at-hand. It is not itself encountered as present-at-hand. The sense of belonging to a world cannot be reduced to an encounter with some object.

> · Hence empirical scientific understanding is limited to a restrictive sense of what *is* and fails to accommodate how we already find ourselves in a world when we engage in scientific inquiry. (Ratcliffe 2012, pp. 144-145)

Furthermore, Wheeler errs by supposing that science gives us access to completely deworlded objects. On the contrary, "presence-at-hand does not escape readiness-to-hand but is a kind of abstraction from it" (Ratcliffe 2012, p. 141). It can be said with certainty that the possibility of abstracting deworlded things presupposes the world: the background in absence of which abstraction does not even makes sense. So the question must be raised: how can empirical results concerned with present-at-hand objects affect the world precisely presupposed in those endeavors? This shows that Wheeler is committed so blindly to the present-at-hand, that his project seems to dissociate readiness-to-hand from the world in which it is intelligible, to then situate it within a scientific (present-at-hand) context, along with adding an interpretation of it in terms of environmental nudges, to finally call it 'Heideggerian' (Wheeler 2012, p. 150). Therefore, the prospects of a Heideggerian cognitive science are criticized by Ratcliffe as a confused undertaking: "the project of seeking to understand our being-in-the-world in cognitive science terms is nonsensical" (2012, p. 138). Two more points can be considered in this regard: (i) Heideggerian topics, when appropriated by a naturalistic cognitive science, remain Heideggerian only in a fairly superficial way (*idem*). And (ii) the latter is confirmed by the fact that the same themes cannot be incorporated into cognitive science without losing sight of Heidegger's philosophy and their place within it (Ratcliffe 2012, p. 139).

What is more: when it comes to reading strategies, it appears even more exact to designate this *sui generis* reception of Heideggerian topics into cognitive science as an 'analytic interpretation' of Heidegger and phenomenology in general. Indeed, as Rehberg (2012) has argued, the interpretation is selective and schematic in what it takes from Heidegger and in what it ignores. So it is not only that analytic-trained philosophers are interpreting Heideggerian themes to incorporate them into a cognitive science discourse—thus overlooking that it was not Heidegger's intention in the first place to give an account of human cognition or to develop a theory of knowledge or a philosophy of mind. It is also that by analytically abstracting some selective topics that belong to the whole fabric of *Sein und Zeit*, the matter of Heidegger's thought is inevitably eradicated. As Rehberg has argued:

> to put aside the issue of Dasein, to translate it into 'human agency' as happens in the service of the new model of cognitive science, or to re-translate it into subjectivity (the very conception from

> which Heidegger seeks to distance his work), is simply to eliminate the meaning of being-in-the-world, whose import in *Being and Time* is precisely to show the derivative nature of any conception of human being in which it is reduced to human agency or (any kind of quasi-Cartesian) subjectivity. (Rehberg 2012, p. 160)

However, based on the assumption that there still may be "multiple ways in which one can naturalize phenomenology, in the sense of integrating phenomenological data, methods, and insights into natural scientific experiments in cognitive science, including psychology and neuroscience, without engaging in naturalistic reductionism" (Gallagher 2012, p. 88), a new project of a phenomenological philosophy of mind (the phenomenological mind project, see Gallagher & Zahavi 2010) is nowadays thriving. So even if phenomenology could be naturalized without engaging into naturalistic reductionism, is that a phenomenological project? What promises would such project hold for the foreseeable future?

The project of a phenomenological philosophy of mind will be resorted to in the next (and last) chapter. This will enable to reach the conclusion of this whole investigation: that philosophy must guard itself against the pressures of confusing its questions with the aims and scope of the sciences.

| 10 |

THE PHENOMENOLOGICAL ALTERNATIVE

Naturalized Phenomenology?

This chapter constitutes the conclusion of this whole investigation into the reception of phenomenology within cognitive science. Consequently, a look shall be taken at the project of a phenomenological philosophy of mind given that it resumes the themes discussed in the previous course, as it also offers the most recent attempt at addressing the hackneyed relationship between phenomenology and science. In addition, this offers the opportunity to characterize cognitive science paradigmatically not as the anthropomorphization of the machine but rather as the mechanization of the human (Dupuy 2009, p. 5): a project that Heidegger was keen to designate as "the technical construction of the human being as machine" (Zoll, p. 178). The irreducibility of philosophical questioning shall, then, be defended and reasons will be given as to why it matters.

Let us just recall that the great French medievalist Etienne Gilson once said that 'philosophy always buries its undertakers' (1999), and it seems that this is crucially put to the test when addressing precisely the theoretical bulk constituting the investigation that comes with this chapter to an end. Thematically, this theoretical bulk comprises the intersection between phenomenology, empirical science, and analytic philosophy of mind in the context of the rise and development of cognitive science with its core project of naturalizing intentionality and artifitializing cognition by technological means.

It seems indeed as though a lot were at stake in the clash of forces between phenomenology and empirical science. As it happens, phenomenology, at the behest of the fine-grained descriptions offered by neuroscience and cognitive psychology, might have to give up its most beloved concepts and descriptions for the sake of scientific pertinence and actuality. It bears just reminding that psychology in Husserl's times was not yet cognitive psychology; computationalism and today's solid information theory were just the initial attempts by cybernetics, which Heidegger then criticized as the new face of metaphysics. So it could be argued that, given the emergence of the new sciences and the technological

advances that led to the network society (Castells 2010a; Castells 2010b; Castells 2010c), or to what more recently Floridi (2014) designates as the 'infosphere,' philosophical discourse must also adapt to the new circumstances. This adjustment implies the acquisition of a new set of metaphors for interpreting centuries old philosophical ideas and theoretical aspirations. In view of this there are pressures that phenomenology should perhaps abandon its original transcendental and antinaturalistic stance. Maybe the latter was just an old (and odd) posture better suited for past scientific disputes which are now beside the point and do not belong to the 'state of the art' of current scientific concerns. So conceivably a transcendental and antinaturalistic stance has now been (or must be) overcome by new scientific endeavors and, above all, by the advent of computer technology which has provided the grounding metaphor for imagining the workings of the mind.

But let us remind ourselves that the discovery of Husserl's "Werk des Durchbruchs"—as he himself called his *Logische Untersuchungen* retrospectively in the preface to the second edition from 1913—is a broadening of intuition: there is more to see than mere objects and mental processes. As is well known, Husserl's investigations open with a lengthy study called 'Prolegomena to Pure Logic', which is famous for Husserl's rather scathing critique of some current modes of thought that are deemed, not inaccurate, but false, absurd, and dangerous, and this inasmuch as they impeded the advancement of reason. But those modes of thought were the mainstream views of the era, and they were currency among the most remarkable philosophers, psychologists, and scientists occupying the most important academic posts in German universities. So the philosophical journey initiated by phenomenological research is marked from the outset by an attempt to struggle against a philosophical idol which purported to replace true philosophical thinking. Husserl, in fact, waged a war against the mainstream views of the time and challenged precisely the professors and scientists holding those very ideas.

The general idea which Husserl confronted and criticized as untenable was not only popular back then, but its realization promised to crown the sort of naturalism espoused by the typical positivism of the time (and my investigation pretends to show that this idea is pervasive today). The idea goes like this: a theory of knowledge is certainly concerned with the nature of thinking, perceiving, judging, believing, and knowing. All faculties which we today would call 'cognitive'. It must be obvious, as it was back then, that these faculties must be investigated by a science fit for them, that is, a science

concerned with the cognitive nature of these faculties and its psychical structure. And this science is, of course, psychology (today we would say cognitive psychology or cognitive science). But it goes much further than this: if it holds true that we arrive at our scientific and logical reasoning by way of the above-mentioned cognitive faculties, then it must also hold that psychology is responsible as well for these laws of thinking. And because psychology is an empirical science, it follows that it must be possible to investigate the laws of thinking from an empirical and naturalistic perspective. Therefore, psychology provides the theoretical foundation for our logical reasoning. After all, we human beings are creatures that evolved from natural environments and our cognitive faculties are to be explained by the very biological circumstances that gave rise to our logical reasoning. This position is refuted by Husserl and its name is psychologism.

Husserl does not deny that there is a possible and necessary science that investigates the nature of cognitive processes as they ensue from neurophysiological strata. I can say in English, for example, that 'Berlin is the capital of Germany' and then say, this time in German, that 'Berlin ist die Bundeshauptstadt der Bundesrepublik Deutschland'. There must be a cognitive neuroscience of language that explains how exactly the brain of a polyglot works. We can for example ask: how does polyglotism affect the brain? One can with all legitimacy ask how the process of second language acquisition restructures the brain and what processes are intervening when someone is competent in several languages. Therefore, Husserl would not deny that this is possible and even necessary. However, the truth of the proposition 'Berlin is the capital of Germany' cannot be found in neural connections. So this is Husserl's refutation of psychologism: that it does not distinguish between the *object of knowledge* (in our case, the truth that 'Berlin is the capital of Germany') and the *act of knowing* (the process taking place when someone affirms that 'Berlin is the capital of Germany'). And this is, according to Husserl, a terrible mistake (a category mistake), because, on closer inspection, it must be apparent that a logical truth is independent of the actual, that is, real (physical) act of carrying out a thought or utter a proposition. So on the one hand, we have the natural event of uttering a proposition, and the neurophysiological process underlying that very act. On the other, however, the meaning of the proposition and its truth validity cannot be said to be subjected to the natural laws underlying the acts of knowing or believing or judging or perceiving or thinking. If this were the case, we would have a

different logic and different mathematical truths, had we evolved differently, which is absurd. Logic is not factual, and in the same vein, it is not affected by the matters of fact. Logic investigates ideal structures and laws, which are characterized by their exactness and certainty. Psychology, in contrast, is an empirical science that investigates the factual nature of thinking, and consequently its results are also characterized by the vagueness and probability of all the other empirical sciences.

Moreover, according to Husserl, the very possibility of knowledge-sharing in different acts of knowing should be argument enough against psychologism. There is a dialectic between the identity of the same meaning and the different acts taking place when communicating that meaning with others. And time and again, Husserl urges his readers to go to the physiological strata to find the truth validity that 'Berlin is the capital of Germany'. Indeed, it cannot be found there. It cannot be found anywhere. Psychologism, for Husserl, entails a self-refuting skepticism: one cannot attempt a naturalistic reduction of ideality to physical reality without undermining the very possibility of any theory, including psychologism itself. So this is what it means to say that psychologism is absurd, for not only it cannot account for a theory of knowledge: it cannot even explain what a theory is. It cannot explain how valid knowledge is communicated between factual, historical beings.

And this is what it means to say that phenomenology is antinaturalistic: that it is absurd to suppose that consciousness is just another object of physical reality that can be investigated as any other object that there is. Consciousness is not as any other object that exists in physical reality. So today's approaches to consciousness which cluster around questions such as: 'how does consciousness emerge from matter?', 'how can we explain consciousness in naturalistic terms?', etc., are psychologistic in the Husserlian sense of the term. Husserl's contribution is then critical (in the Kantian sense of 'Kritik'): because it establishes limits to what can and cannot be investigated from the empirical perspective.

From the phenomenological standpoint, consciousness is precisely not amenable to the objective perspective that characterizes empirical research, because it deals also with structures and laws that are not empirical in nature. This is why the character of phenomenology as a research program is to be understood as a fundamental coupling of meaningful acts of knowing and the objective (logical) contents that are not to be confused with the former acts. And this is very important to remember:

phenomenology is not only interested in the first-person perspective, as is usual to say today. Although mental contents are meaningful because they happen to someone, they make their appearance to someone, and are perceived by someone, Husserl's point is that even if they were not perceived or lived, a truth validity would not be affected in the least (it would not be affected even if we did not exist). As Husserl asserts in *Phänomenologische Psychologie,* no truth is a fact: "the truth that 2+3=5 stands all by itself as a pure truth whether there is a world, and this world with these actual things, or not" (Hua IX, p. 23). Moreover, "purely ideal truths are 'a priori' and are discerned as truths in the unconditioned necessity of their universality" (*idem*).

Now, this is not to say that phenomenology's anti-naturalistic stance implies at the same time a rejection of natural science *per se*. It would be useful to recall Husserl's own words in *Ideas* (1913): "When natural science really speaks, we listen with joy and as disciples. But it is not the case that natural science always speaks, when natural researchers do so, and certainly not when they speak of natural philosophy and of a naturalistic theory of knowledge" (Hua III/1, 45). This is indeed the case when they confuse the *object of knowledge* (which is ideal) with the *act of knowing* (which is concrete, historical, and subjective). This position commits the error of ignoring the fundamental difference that exists between the domain of logic and psychology, because logic (as well as mathematics and formal ontology) is not an empirical science concerned with factually existing objects.

However, in the discussion concerning the relation between phenomenological philosophy and cognitive science, the abandonment of the transcendental standpoint is meant to serve as a prerequisite for phenomenology's suitability to scientific standards. Scientific advances brought about by the 'cognitive turn' since the time of Husserl and Heidegger might make it conceivably possible to overcome this antinaturalism (see Petitot, Varela, Pachoud & Roy [eds.] [1999], pp. 39-43). And this is why the project of 'naturalizing phenomenology' must be dealt with, since precisely the 'seductions of naturalism' (Crowell 2012) have been lurking below the surface of the cognitivist project. Indeed, at the heart of cognitivism—which was defined in the introduction as the 'philosophy behind cognitive science'—lies the program of dealing with the objects presented by intentionality, whereby 'dealing' with them amounts either to reduce them to smaller natural components or simply to eliminate them. This is, as Husserl denounced, "a consistent elimination of all 'merely subjective' properties belonging

to the things of immediate experience, of all features stemming from subjectivity" (Hua IX, p. 54); a methodological decision of the natural sciences, whose paradigm of maximum abstraction is physics, that has permeated psychology and the philosophy of mind alike.

A general observation that might ensue from the current state concerning the relationship between philosophy and empirical science—or, as in our case, between phenomenology and cognitive science—can be expressed straightforwardly as follows: cognitive science has reached so hegemonic a place within present-day scientific discourse and praxis, that philosophy seems to feel obliged to show its propinquity to its methods and demands. So it is a two-way avenue: empirical science would devour intentionality if it could even at the price of getting rid of it, while philosophy feels tempted to explore the possibilities of turning its objects into natural phenomena. In an epilogue to his *The Mind's New Science*, Gardner is pleased to report that the reputation and scope of cognitive science "have risen more rapidly than I would have predicted" (1987, p. 398). Given that the hegemonic place of cognitive science is no less reputable today than it was almost thirty years ago—indeed, judging from funding opportunities, the continuing opening of graduate programs, and Hollywood films,[101] the prestige of cognitive science can be said to be today at its highest—philosophy seems to need to exhibit its scientific pedigree, that is, its competence and proficiency for enriching the cognitive landscape. Proceeding otherwise would undermine its current credibility.

For much the same reason, the aforementioned contemporary predicament of philosophy of being mere *ancilla scientiarum*—or, in more Quinian terms, there being no place for a prior philosophy over against the empirical spirit that animates natural science (Quine 1969, p. 26)—raises the need for demarcation[102] vis-à-vis the hegemonic scientific discourse. Here I just want to stress that this, incidentally, has been always the point of tension in the last chapters of my investigation: the demarcation as to how philosophy stands in relation to the pressures of hegemonic empirical science

[101] Three recent examples of Hollywood movies with a cognitive background are *Her* (2013) directed by Spike Jonze, *Robocop* (2014) directed by José Padilha, and *Transcendence* (2014) directed by Wally Pfister. In the first film, the main protagonist falls in love with his new purchased intelligent operating system. On the other hand, *Robocop* is not only an example of the cultural hype attained by cognitive science but also of philosophy making it to the mainstream: the roboticist is named Dennett, the senator who is against the proliferation of androids in society is named Dreyfus, and OmniCorp's CEO is featured by a character named Sellars. In *Transcendence*, the technological singularity about which Kurzweil and Chalmers dream with passion is finally attained, although it turned out to be a nightmare.

[102] See, on this regard, Lembeck's so-called 'Abgrenzungsproblem' (2010, p. 139 ff).

(in our particular case, of cognitive science.) In contrast, other ways, that is, possible measures for integration between discourses that might appear as nemeses—but need not be—are certainly in the offing. In either case, my conclusion is that the result of this tension has been philosophy's capitulation to the enormous weight of the cognitive. The cognitive is so pervasive that nowadays a transcendental role is ascribed to it: "what makes science possible? Specifically, what features of the human mind, of human cognitive development and of human social arrangements permit and facilitate the conduct of science?" (Carruthers, Stich & Siegal 2002, p. 1). Cognitive science is not just one among many other theoretical endeavors: science itself in general must have a cognitive basis, for knowledge is possible only on the basis of cognition. A perfect circle.

The conclusion to which I have arrived regarding the state of the art of the discussion that has been presented throughout this investigation can be expressed as follows: every single researcher enthused about the prospects of putting to work phenomenological insights to underwrite the importance and current scientific validity of phenomenology—every single one of them, from Winograd to Dreyfus, through Agre and Wheeler—has succumbed to a species of naturalism that certainly deforms the phenomenological enterprise. This rendition to naturalism was confirmed in two ways throughout the story that was narrated in the previous chapters. On the one hand, the first section of this investigation presented researchers who implied their technological toys would finally realize what philosophy could not for several centuries. Mind—so the claim went—could be finally mechanized without remainder. In turn, the second section dealt with researchers who appraise what phenomenology might contribute to the current state of cognitive science, especially to help it escape from the Cartesian mindset pervading traditional cognitive science. It was a matter, then, not of debunking cognitive science *tout court* but of refining (and redefining) it. These researchers, however, conceived of this contribution as one that modifies empirical science providing it with new avenues to traverse, but this on pain of modifying phenomenology itself, of naturalizing it to a certain extent. Therefore, in order to complete the story, it is needed to render an account as to what extent this naturalization of phenomenology is possible without trespassing the frontiers between empirical research and philosophical questioning. More generally, what are the consequences of conflating the task of philosophy and the endeavors of empirical science?

The newest kid in the neighborhood, which is more phenomenologically informed than the projects of, say, Dreyfus, Agre, and Wheeler, is the project of a phenomenological philosophy of mind: the phenomenological mind, represented by Gallagher and Zahavi (2010)—but also espoused by a plethora of researchers, from 4E researchers, neurophenomenologists, to enactivist neuroscentists (McClamrock 1995; Petitot, Varela, Pachoud & Roy [eds.] 1999; Lakoff & Johnson 1999; Noë 2006; Thompson 2007; Petit & Berthoz 2008; Chemero 2009; Steward, Gapenne & Di Paolo [eds.] 2010; Hutto & Myin 2013).[103] The devil—as the saying goes—is in the details, so it needs scarcely be said that a thorough account of the aforementioned authors and their approaches would deserve a larger treatment than is here possible to offer.[104] However, it should be possible to tackle what is essential in this approach: the insistence in the need for 'naturalizing phenomenology,' either by a straightforward formalization (even a mathematization, see Petitot, Varela, Pachoud & Roy [eds.] [1999], p. 42; Marbach 1993) of some phenomenological concepts and descriptions, or by means of a felicitous integration in which both, phenomenology and empirical science, learn from each other; an integrative approach that effectuates also a translation of phenomenological concepts like 'Dasein' into more cognitive constructs like 'agency,' and where 'being-in-the-world' is turned into 'activity.' Thus, phenomenology and cognitive science can be thought of as mutually constraining: "it is quite possible that the mutual constraint situation will lead to a productive mutual enlightenment, where progress in the cognitive sciences will motivate a more finely detailed phenomenological description developed under the regime of phenomenological reduction, and a more detailed phenomenology will contribute to defining an empirical research program" (Gallagher & Varela 2001, p. 21). This peaceful deal was also suggested by the neuroscientist Jean-Pierre Changeaux in conversation with Paul Ricoeur, when he declared that his purpose was not going to war with phenomenology: "to the contrary, [he wanted] to see what constructive contribution it can make to our knowledge of the psyche, acting in concert with the

[103] Researchers attached to the aims of the phenomenological mind project, gravitate incidentally around the journal founded for these purposes: *Phenomenology and the Cognitive Sciences*, edited by Sean Gallagher and Dan Zahavi, and published by Springer. The journal presents itself as a forum for illuminating the intersections between phenomenology, empirical science, and analytic philosophy of mind, that builds bridges between the Husserlian tradition and other disciplines and addresses recent work on the connection between empirical results in experimental science and the first-person perspective.

[104] Ebinger (2012), for instance, has dedicated a whole investigation to the transformation of phenomenology through Varela's neurophenomenology.

neurosciences" (quoted by Gallagher & Varela, p. 18). The latter also seems in accord with Zahavi's understanding of the naturalization of phenomenology:

> to naturalize phenomenology might simply be a question of letting phenomenology engage in a fruitful exchange and collaboration with empirical science. Phenomenology does study phenomena that are part of nature and therefore also open to empirical investigation, and insofar as phenomenology concerns itself with such phenomena it should be informed by the best available scientific knowledge. (2010, p. 8)

So what is the guiding idea behind the project of naturalizing phenomenology? According to Gallagher and Varela (2001, p. 18), it is actually but an irony that it is the scientists in general the ones seeking more readily an integration of phenomenology into a naturalistic account of cognition. In contrast, phenomenologists "have been satisfied in drawing critical lines that identify its limitations [that is, of cognitive science]" (p. 17). However, it is my conviction that pointing out these limitations is not just bad faith, empty criticism, outright lack of scientific enthusiasm, or worse a neo-Luddism opposing vehemently modern forms of technology and empirical research. Instead, opposing the cognitivization of philosophy serves the *antiscientistic* (not antiscientific) purpose of making clear that the cognitive story is not all there is, and that something important will be forever lost if one confuses the themes of philosophy with those of empirical science.

Regarding naturalization projects, let us note that they have emerged due to the fact that naturalism is pervasive and hegemonic in the scientific Zeitgeist. Indeed, within contemporary philosophy, naturalism—or some sort of it—is the default metaphysical position. As Roy Wood Sellars famously claimed almost a century ago in his *Evolutionary Naturalism* (1922), "we are all naturalists now"—or perhaps must be. Not being a naturalist these days can cause one many problems, among them not being taken seriously. Remember Wheeler's (2005) Muggle constraint which appeared in the previous chapter: in the investigation of everything there is, science is masterfully in charge and philosophy must simply back off if it arrives at conflicting statements with established scientific theories. 'Spooky' stuff must be avoided at all costs in whatever theories are postulated. However, everything depends on what it means to abide by this constraint that philosophy does not overstep the mark delineated by empirical science and, in general, on understanding what the project of naturalization amounts to in more general lines.

In the first place, one of the first impulses of contemporary philosophical naturalism was the demand to do something with all those nonscientific, albeit psychological-like, ascriptions populating the commonsense framework. Human relationships and meaningful interactions where interpersonal understanding occurs are said to be traversed by means of a commonsense or folk psychology. Paul Churchland defined folk psychology as the "prescientific, commonsense conceptual framework that all normally socialized human beings deploy in order to comprehend, predict, explain, and manipulate the behavior of humans and the higher animals" (1998, p. 3). More crucially, Churchland has it that from the point of view of eliminative materialism, the thesis which he espouses concerning the folk psychological framework, "our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather smoothly reduced, by completed neuroscience" (1981, p. 601).

In this respect, Churchland is a representative of the theory-theory of mind, according to which every aspect of human experiential affairs is funded is some sort of theory, so that even emotions, perceptions, and intentions are somehow theoretically postulated entities. On this view, folk psychology—albeit false—constitutes a theory that postulates a wrong picture of the world. A developmental psychology based on the assumptions of the theory-theory of mind holds that this theoretical capacity of having a folk psychological theory is learned as the child grows up (Churchland 1991). So growing up as a human child is tantamount to developing a theory of mind and maturing cognitive competences amounts to engaging "in an inferential leap beyond what we can actually observe about other people—their 'behavior'—in order to relate to them truly as psychological beings with, for example, intentions, emotions and thoughts" (Costall & Leudar 2009, p. 40). Philosophy of mind must then tackle how humans develop such theory and offer scientific corrections to the mistaken assumptions pervading this folk theory. If one is tuned to this idea, it is only natural to conclude that the folk psychological framework is "a stagnant or degenerating research program, and has been for millennia" (Churchland 1981, p. 75). From a certain point of view, then, is not at all exaggerated to call this madness (Costall & Leudar 2009), for it is suggested that the way human beings learn to cope with their world is a degenerating research program! So the project of naturalizing

intentionality, of rendering language and thought as just biological categories (Millikan 2001), is meant to avoid what Fodor called the permanent stubbornness of semantics to be suitably integrated in the natural order (1984, p. 232). Of course, this is rather an extreme reductive version of naturalism, but it remains an influential view. While AI researchers were trying to design programs with common sense (McCarthy 1958), eliminative materialism looked for establishing a view according to which common sense did not even deserve the effort to be programmed, since it was really a defective theory. From the perspective of eliminative materialism classical AI, or GOFAI, failed in its attempts to program intelligence. What was needed was instead a scientific theory stripped of commonsense ascriptions.

So against the backdrop of a radical eliminative materialism of the kind just depicted, newer projects emerged which sought to develop a dialogue between phenomenology and nonreductionist cognitive science (Gallagher 2010). These new ventures fight eliminative materialism not only because of its ruthless reductionism (Bickle 2003), but above all because it is utterly false and does not remain true to the facts. As Gallagher (2010) argues: "my experience of my action is not purely a motor experience—it is also world-involving, since it is directed at some task in the world" (p. 28). The point is that experience cannot be simply reduced to the neural underpinnings:

> the use of phenomenology in the empirical cognitive sciences reinforces the importance of first-person experience and thereby undermines the reductionist tendencies that one often finds in scientific theory. Since we are more than just a bunch of neurons, one requires good methods that will allow us to sort out what the '*more*' is. Phenomenological enlightened ways of understanding first-person experience and how it affects experimental results just is one such method (Gallagher 2010, p. 32)

Agreeing with this view, Zahavi (2010) proposes a modest adaptation of phenomenology into the naturalistic framework, engaging in a fruitful exchange and collaboration with empirical methods. Incidentally, neither Gallagher nor Zahavi expect empirical science to swallow up phenomenology in its entirety. For Gallagher, the new scientific use of phenomenology does not challenge completely or annul classical phenomenology (2010, p. 32). Furthermore, on Zahavi's account, "the ultimate aim of phenomenology is to provide a transcendental philosophical clarification, and as such its aim differs from that of empirical science" (2010, p. 8). There is a nonreducible core of phenomenology which is not positive and which, for that very reason, cannot be objectified, since "to claim that phenomenological content is positive is to claim that it is derivative upon something more basic" (Rowlands 2010b, p. 201). And inasmuch as nonpositive, this transcendental stance brought

forth by phenomenology is also noncognitive: "the transcendental mode of presentation is that aspect of the experience *in virtue of which* the experience's object is presented as falling under, or possessing, a given aspect or empirical mode of presentation… The noneliminable, transcendental core of the experience, therefore, consists in a form of *revealing* or *disclosing activity*" (Rowlands 2010b, p. 189).

How is it, then, that notwithstanding this noneliminable core of the transcendental, phenomenology can be said to come under naturalist scrutiny on pain of modifying itself in the process? For Gallagher, the possibility of moving to and fro, from phenomenology to empirical science and then back again, was suggested by Husserl himself in a much quoted paragraph of *Cartesianische Meditationen*: "every analysis or theory of transcendental phenomenology… can be realized in the natural realm [*auf dem natürlichen Boden*], by giving up the transcendental attitude" (Hua I § 57, p. 159). This is Husserl's way of saying that this transcendental naïveté allows for the appearance of a purely internal psychological theory ("*eine 'innenpsychologische' Theorie*"). Crucially, this is a psychology, according to Husserl, that—"although it merely explicates what belongs to a psyche, to a concrete human Ego, as its own intentional essence"—corresponds to a transcendental phenomenology, and vice versa (*idem*).

According to Zahavi, it must be recognized that "phenomenology deals with topics that it shares with other disciplines, and it would be wrong to insist that it should simply ignore empirical findings pertaining to these very topics" (2010, p. 8). On Zahavi's view, this is no invention of current philosophical discourse. Rather, the integrative approach has antecedents in Husserl himself: "on the one hand, we have *transcendental phenomenology*, and on the other, we have what he [Husserl] calls *phenomenological psychology*" (Zahavi 2010, p. 10). The former, explains Zahavi, has a more philosophical vocation and remains strictly bounded to the transcendental stance interested in the constitutive dimension of subjectivity. The latter is a sort of philosophical psychology which takes the first-person perspective seriously but remains within the natural attitude: "the difference between the two is consequently that phenomenological psychology might be described as a regional-ontological analysis which investigates consciousness for its own sake" (Zahavi, *idem*). Then Zahavi, as Gallagher above, quotes § 57 from *Cartesianische Meditationen* and concludes that Husserl "spoke of a parallelism between phenomenological psychology and transcendental phenomenology and claimed that it is possible to step from one to the other through an attitudinal change" (2010, p. 11). Moreover, "might such

considerations allow for the possibility that empirical findings (if based on meticulous analysis of the phenomena and if subjected to the requisite modifications) could be taken up by, and consequently influence or constrain the analysis of transcendental subjectivity? I see no reason why not" (Zahavi, *idem*).

However, I still do not see how exactly such empirical findings could affect the transcendental stance. And I do not think Zahavi is clear on it either. Quoting Tatossian's (2002) critique of Merleau-Ponty's speculative use of empirical findings, Zahavi maintains that, when it comes to empirical science, "one shouldn't remain in the ivory tower of the transcendental philosopher" (2010, p. 12 n. 12). But Husserl, to be sure, considered such influence and constraints of empirical findings over the transcendental stance as absurd as if one expected the essence of numbers to be affected by laboratory experiments. So the mutual-constraint principle, at least from a Husserlian point of view, is highly contentious.

Gallagher and Zahavi (Gallagher 2010; Gallagher 2012; Gallagher & Zahavi 2010; Zahavi 2010) offer several examples of the use of empirical findings providing first-person descriptions that, on their view, might be of utmost relevance for phenomenology. Among them are neuropsychological descriptions of anosognosic disorders of body-awareness, psychopathological descriptions of schizophrenic disturbances of self-experience and intentionality, developmental descriptions of social interactions in early childhood, ethnological descriptions of culture specific emotions, etc. According to Gallagher (2010), phenomenological distinctions like, for instance, the ones between a sense of agency and a sense of ownership can—via empirical work—be traced back to the neuronal processes that generate the first-order phenomenal experience. In such a way, "experimental science can assist phenomenology in drawing out the complexity of such phenomenological distinctions" (p. 28). Zahavi argues that, although phenomenology does not address the subpersonal mechanisms that enable experience, psychopathology, developmental psychology, cognitive psychology, anthropology, etc. "can provide person-level *descriptions* that might be of phenomenological relevance" (2010, p. 8).

However, it is not always clear how these empirical findings modify or affect what is prior to any scientific theorizing and that renders this theorizing possible! I agree with Zahavi that one should be informed by the best available scientific knowledge (2010, p. 14), but this is not something radically

new. All great philosophers were interested in the scientific endeavors pervading their time, and Husserl himself was educated as a mathematician. But with regard to this way of incorporating phenomenology in cognitive science infusing the phenomenological mind project, it seems rather as though the word phenomenology were on occasion used in a nonphenomenological way. Indeed, the term 'phenomenology' seems to be conflated to mean both the what-is-likeness of subjective lived experience, but also phenomenological philosophy. For example, when Gallagher refers—in a recent study—to "The Cruel and Unusual Phenomenology of Solitary Confinement" (Gallagher 2014), he is not using 'phenomenology' with its transcendental potential, but rather as the first-person experience of undergoing solitary confinement. Furthermore, this is also the usage of 'phenomenology' as a qualitative method in psychological practice according to which the focus should be placed on people's perceptions and lived experience. Phenomenology is thus understood as an accurate description of subjective experience (Langdridge 2007). However, to say it with Husserl: "for without having seized upon the peculiar ownness of the transcendental attitude and having actually appropriated the pure phenomenological basis, one may of course use the word phenomenology; but one does not have the matter [*die Sache*] itself" (Hua III § 87, p. 179). So it all turns out to be about the *Sache*, the matter itself or the business of philosophy. But is this *Sache* just the subjective character of the first-person perspective? This is indeed what needs to be clarified.

In the following section, I present a coda. I argue that the problem with the naturalization of phenomenology is that it conflates the *Objekte* of empirical science and the *Sache* of philosophy, and that doing so has always been present in the career of the cognitive scientific endeavors and continues to be a threat in the recent reception of phenomenology in cognitive science.

Coda: The Irreducibility of Philosophical Questioning

The question still remains: if there is a nonreducible and noncognitive core with which phenomenology concerns itself, what does this very fact say about the different tasks at hand pertaining respectively empirical science and philosophy? The existence of a nonreductive core which is safeguarded by phenomenology does not only demonstrate that consciousness is irreducible to whatever physical

processes but, more importantly, that philosophical questioning is nonisomorphic with the theoretical attitude of the natural sciences.

The problem of the latter approaches seeking a close connection with empirical findings is precisely that this nonisomorphism, the crucial and unique place that is to be conferred to philosophy alone, might go astray when confronted with the seductions of naturalism. For it is the assumption of naturalism what led Wheeler, for instance, to warn philosophers of not postulating occult entities. However, the critical question must be asked: is this really what philosophy is about? It would be interesting to note—and I say this with outright irony—which entities and natural processes have been introduced by phenomenologists that are inconsistent with empirical science! We can be sure at least of this: Husserl did not offer a neuroscientific theory of phenomenality and Heidegger's cognitive psychology is nowhere to be found. So it is rather striking that someone would suggest that one should beware of the theoretical objects being postulated by phenomenology, when certainly neither Husserl nor Heidegger were in the business of populating their 'theories' with objects challenging those processes investigated by empirical science. Neither Husserl nor Heidegger presented scientific investigations. So let us be clear at this juncture: there is no phenomenological theory of mind, if by 'mind' one understands an ontic region, whose law-like physical processes have to be addressed and explained by means of empirical methods. If we are not aware of Heidegger's theory concerning the neural correlates of consciousness or Husserl's ideas on how to design a quantum computer, the reason is because they do not present us with any kind of theory regarding natural (not even artificial) phenomena. Phenomenology does not postulate entities whatsoever, neither natural nor cultural, and it certainly does not concern itself with the law-like causality governing natural processes. Instead, its task is to investigate the very basis that makes it possible for science to construct theories populated by such objective invariances and the very constitution of the experience which allows for those objects to appear as meaningful. Following Husserl, this transcendental field that is revealed by phenomenology also discloses a transcendental subjectivity that cannot simply be identified with any constituted entity, whether natural or cultural. The transcendental sphere is, then, always relied on when postulating theories about consciousness or the mind, "while not realizing that we are doing so, since whether

acknowledged or not, it is the unspoken concomitant of everything we think and believe" (Olafson 1987, p. 256).

Phenomenology concerns itself not only with the *noema*, or the ways that the object appears in experiences of various sorts but crucially also with the *noesis*, or the experience through which the self-showing object becomes manifest (Hua III §§ 87-96). Moreover, phenomenology is crucially *noema-noesis* correlation research and as such it is certainly guided by a transcendental subject-oriented interest, but its explanatory attempts are not exhausted by subjective experience (see Lembeck 1999). Phenomenology is not only about how things appear to *me* but more importantly about the structure of appearance itself. Therefore, phenomenology investigates reality, not just lived experience. So phenomenology should not be mistakenly conceived solely as the first-person perspective (not from Dennett's perspective, to be sure, but neither from Varela's, nor Zahavi's, nor Gallagher's). In this respect, allow me to add that it is actually strange that the phenomenological mind project would pigeonhole phenomenological research as pertaining solely to the first-person perspective. And this confusion is both shared by critics of phenomenology, like Dennett, and researchers fond of it, like Gallagher and Zahavi. For the former, phenomenology, that is, first-person ascription-making, is illusory. For the latter researchers, phenomenology is also about the first-person perspective, but they regard it with high esteem. Both positions might seem contradictory, but they coincide in that phenomenology is simply regarded as the what-it-is-likeness of experience. But let there be no mistake: both critics and first-person perspective enthusiasts coincide in that the phenomenological garden is populated somehow by first-person events.

However, the point at issue brought forth by phenomenological philosophy with its antinaturalistic transcendental stance is neither contested nor slightly affected by the purported mechanical organization of the intentional, which Dennett, for instance, imagines as demolishing the phenomenological edifice. Nor is it refined—and this against the phenomenological mind project—by more accurate descriptions of first-person data under the auspices of empirical experiments guided by phenomenological concerns. The discussion can be framed along two mutually excluding propositions: either (i) there is phenomenology (in the sense of first-person data) or (ii) there isn't. For the second proposition to be true, it is necessary to argue that the absence of phenomenology amounts to the fact

that first-person data are available or are translatable to third-person data. One of the most entrenched convictions of cognitivism is precisely that this translation is possible. Were it not possible, then cognitive science would be at the peril of not being able to explain subjective experience. The defenders of the first proposition (Nagel, Varela, Zahavi, and Gallagher, for instance) simply need to show that the availability of first-person data is only possible from the first-person perspective. And that this very fact is important.

What is crucial to understand, however, is that phenomenological philosophy is not in the business of obtusely defending the existence of those experiences populating the 'phenomenological garden,' whose spell Dennett thinks to have broken. This is the same personal point of view that the phenomenological mind theorists strive to ensure by means of an empirically informed phenomenology. According to this appraisal of the first-person perspective, the spell is richly real and fundamentally important. Dennett's associates with what he calls the standard denizens of the phenomenological garden (1991, p. 45) experiences such as pains and smells, daydreams, images of fancy, and sudden intuitions, as well as emotive experiences like pride, anger, or remorse. His point is rather straightforward: if consciousness is, in fact, populated by those phenomena, then consciousness is reducible to objectively expressed information that can be obtained from third-person investigations. But what Dennett ignores is that Descartes would easily agree with him that such mental states and subjective experiences are not indubitable. As Descartes argued in the second meditation on first philosophy, even all those experiences and mental contents which we wholeheartedly adhere to and find indubitable, could actually be a product of an hallucination or a dream. On the contrary, what is indubitable is the *act of attending* to these contents, the vertical experience of living the sense of those contents. And what Descartes wants to say with this, as Sáez (2006) has masterly argued, is that "what properly belongs to the sphere of the subject is the dimension of self-apprehension *in actu*. Everything that Dennett includes as component of the phenomenological garden has the form of a *content that is experienced*. In no way does he refer to the *acts* themselves by means of which one attends to these contents" (p. 122). So appropriately understood, the trascendental stance characterizing a phenomenological act of sense is not affected neither by breaking the spell of the phenomenological

garden, nor by refining experimentally the contents of subjective experience. From this point of view, both critics as well as first-person enthusiasts are mistaken.

Another way of explaining this irreducibility of philosophy to science can be also couched in terms of Heidegger's question concerning the meaning of Being:

> Suppose that there were no indeterminate meaning of Being, and that we did not understand what this meaning signifies. Then what? Would there just be one noun and one verb less in our language? No. *Then there would be no language at all*. Beings *as such* would no longer open themselves up in words at all; they could no longer be addressed and discussed. For saying beings as such involves understanding beings as beings—that is, their Being—in advance. Presuming that we did not understand Being at all, presuming that the word 'Being' did not even have that evanescent meaning, then there would not be any single word at all. We ourselves would never be those who *say*. We would never be able to be those who we are. For to be human means to be a sayer. Human beings are yes- and no-sayers only because they are, in the ground of their essence, sayers, *the* sayers. That is their distinction and so also their predicament. It distinguishes them from stone, plant, and animal, but also from the gods. Even if we had a thousand eyes and a thousand ears, a thousand hands and many other senses and organs, if our essence did not stand within the power of language, then all beings would remain closed off to us—the beings that we ourselves are, no less than the beings that we are not. (GA 40, pp. 62-63)

On Heidegger's view, this fact—that we understand Being—has the highest rank (*den höchsten Rang*), inasmuch as in it "a power announces itself in which the very possibility of the essence of our Dasein is grounded" (GA 40, p. 63). And he then adds: "it is not one fact among others, but that which merits the highest worth according to its rank [*gemäß seinem Rang die höchste Würdigung*], provided that our Dasein, which is always a historical Dasein, does not remain a matter of indifference to us" (*idem*).

This is precisely what is meant by the expression 'the primacy of phenomenology over cognitivism:' a primacy must be given to phenomenology in the investigation of this fact of highest rank, since phenomenology has concerned itself with it and has put it forth as the theme of philosophical questioning. On the contrary, the cognitivist story told in this book has showed how *Seinsverständnis* has rather been reduced, simplified, held back and postponed for later, as though it were just a missing ingredient that could be added to the recipe of cognition. As such, *Seinsverständnis* has remained a matter of indifference to cognitivism.

Philosophy, on the other hand, remains a tragic endeavor inasmuch as it concerns itself with questions that have no definitive answer, that is, philosophy finds it important to devote itself to phenomena that cannot be reduced to the objective third-person perspective. The tragedy of empirical-informed philosophy, to paraphrase Grondin (2000), is that it can contribute to depriving philosophy of its topics and its questions. But philosophy must remain faithful to these questions that cannot be

solved by science, "but that will exist as long as mortals are confronted with the challenge of their finitude" (Grondin 2000, p. 83).

Bibliography

Abramiuk, M. (2012). *Foundations of Cognitive Archaeology*. Cambridge, MA/London: The MIT Press.

Adams, F. & Kenneth A. (2009). "Why the Mind is Still in the Head." In: Robbins, P. & Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*. Cambridge/New York: Cambridge University Press, pp. 78-95.

_____. (2010). "Defending the Bounds of Cognition." In: Menary, R. (ed.) *The Extended Mind*. Cambridge, MA/London: The MIT Press, pp. 66-80.

Adorno, T. (1964). *Jargon der Eigentlichkeit. Zur deutschen Ideologie*. Frankfurt am Main: Suhrkamp.

Agamben, G. (2002). *Das Offene. Der Mensch und das Tier*. Frankfurt am Main: Suhrkamp.

Agrawala, A. & Sam N. (1992). "Computer Systems." In: Shapiro, S. (ed.) *Encyclopedia of Artificial Intelligence*. New York: John Willey & Sons, pp. 241-248.

Agre, P. (1988). *The Dynamic Structure of Everyday Life*. MIT PhD Thesis (also MIT AI Laboratory Technical Report 1085).

_____. (1996). "Computational Research on Interaction and Agency." In: Agre, P. & Rosenscheim, S. (eds.) *Computational Theories of Interaction and Agency*. Cambridge, MA/London: The MIT Press, pp. 1-52.

_____. (1997). *Computation and Human Experience*. Cambridge, MA: Cambridge University Press.

———-. (2002). "The Practical Logic of Computer Work." In: Scheutz, M. (ed.) *Computationalism. New Directions*. Cambridge, MA/London: The MIT Press.

_____. (2005). "The Soul Gained and Lost: Artificial Intelligence as a Philosophical Project." In: Franchi, S. & Güzeldere, G. (eds.) *Mechanical Bodies, Computational Minds. Artificial Intelligence from Automata to Cyborgs*. Cambridge, MA/London: The MIT Press.

Agre, P. & Chapman, D. (1987). "Pengi: An Implementation of a Theory of Activity." *Proceedings of the Sixth Annual Meeting of the American Association of Artificial Intelligence*. Seattle: Morgan Kaufmann, pp. 268-272.

_____. (1988). "What are plans for?" MIT AI Laboratory Memo 1050a.

Agre, P. & Horswill, I. (1997). "Lifeworld Analysis." *Journal of Artificial Intelligence Research*. Vol. 6, No. 1, pp. 111-145.

Aizawa, K. (1992). "Connectionism and Artificial Intelligence: History and Philosophical Interpretation." *Journal for Experimental and Theoretical Artificial Intelligence*. Vol. 4, No. 4, pp. 295-313.

Anderson, D. et al. (2008). *An Introduction to Management Science. Qualitative Approaches to Decision Making*. Ohio: Thomson South-Western.

Andler, D. (2000). "Context and Background: Dreyfus and Cognitive Science." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping and Cognitive Science. Essays in Honor of Hubert L. Dreyfus*. (Volume 2.) Cambridge, MA/London: The MIT Press, pp. 137-159.

Anscombe, G. (2000). *Intention*. Cambridge, MA: Harvard University Press.

Antony, Louise (2002) "How to Play the Flute. A Commentary on Dreyfus' Intelligence Without Representation." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 395-401.

Arbib, M. & Hesse, M. (1987). *The Construction of Reality*. Cambridge/New York: Cambridge University Press.

Bach, J. (2009). *Principles of Synthetic Intelligence. PSI: An Architecture of Motivated Cognition*. Oxford/New York: Oxford University Press.

Baeumler, A. (1967). *Das Irrationalismusproblem in der Ästhetik und Logik des 18. Jahrhunderts bis zur Kritik der Urteilskraft*. Halle: Niemeyer.

Barret, J. (2011). *Cognitive Science, Religion, and Theology. From Human Minds to Divine Minds*. Pennsylvania: Templeton Press.

Barwise, J. & Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA/London: The MIT Press.

Bayne, T. & Montague, M. (2011) "Cognitive Phenomenology: An Introduction." In: Bayne, T. & Montague, M. (eds.) *Cognitive Phenomenology*. Oxford/New York: Oxford University Press, pp. 1-34.

Bechtel, W. (1988). *Philosophy of Science. An Overview for Cognitive Science*. New York: Psychology Press.

Bechtel, W. & Graham, G. (eds.) (2008). *A Companion to Cognitive Science*. Oxford: Blackwell.

Bennett, M. & Hacker, P. (2003) *Philosophical Foundations of Neuroscience*. Malden, MA/Oxford: Blackwell.

_____. (2013). *History of Cognitive Neuroscience*. Malden, MA/Oxford: Wiley-Blackwell.

Bickle, J. (2003). *Philosophy and Neuroscience. A Ruthlessly Reductive Account*. Dordrecht/Boston/London: Kluwer.

Bickle, J. & Ellis, R. (2005). "Phenomenology and Cortical Microstimulation." In: Smith, D. W. & Thomasson, A. (eds.) *Phenomenology and Philosophy of Mind*. Oxford/New York: Clarendon Press/ Oxford University Press.

Blackburn, S. (2005). *The Oxford Dictionary of Philosophy*. Oxford/NewYork: Oxford University Press.

Blackmore, S. (2005). *Consciousness. A Very Short Introduction*. Oxford/New York: Oxford University Press.

_____. (2010). *Consciousness. An Introduction*. New York: Routledge.

Block, N. (1995). "On a Confusion about a Function of Consciousness." *Behavioral and Brain Sciences*. 18, pp. 227-287.

Blumenberg, H. (2009). *Wirklichkeiten in denen wir Leben: Aufsätze und eine Rede*. Stuttgart: Philipp Reclam.

_____. (2010). *Paradigms for a Metaphorology*. Trans. by R. Savage. New York: Cornell University Press.

Boden, M. ed. (1990). *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press.

Boden, M. (1995). "AI's Half Century". *AI Magazine*. Vol. 16, No. 4, pp. 96-99.

_____. (1998). *Computer Models of the Mind. Computational Approaches in Theoretical Psychology*. Cambridge: Cambridge University Press.

_____. (2006). *Mind as Machine. A History of Cognitive Science*. Vols. I & II. Oxford: Clarendon Press.

Bostrom, N. (2014). *Superintelligence. Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bourdieu, P. (1988). *L'Ontologie Politique de Martin Heidegger*. Paris: Les Éditions de Minuit.

_____. (1990). *The Logic of Practice*. Trans. by R. Nice. Stanford, CA: Stanford University Press.

_____. (1991). *La Ontología Política de Martin Heidegger*. Trans. C. de la Meza. Barcelona/Buenos Aires/ México: Paidós.

Boyer, P. (ed.) (2008). *Cognitive Aspects of Religious Symbolism*. Cambridge/New York: Cambridge University Press.

Bracher, M. (2013). *Literature and Social Justice. Protest Novels, Cognitive Politics, and Schema Criticism*. Austin, TX: University of Texas Press.

Braver, L. (2013). "Never Mind. Thinking of Subjectivity in the Dreyfus-McDowell Debate." In: Schear, J. (ed.) *Mind, Reason, and Being-in-the-World. The McDowell-Dreyfus Debate*. London/New York: Routledge, pp. 143-162.

Bremermann, H.-J. (1962). "Optimization Through Evolution and Recombination". In: Yovitts, M. C. et al. (eds.) *Self-Organizing Systems*. Washington D. C.: Spartan Books, pp. 93-106.

Brooks, R. (1991). "Integrated Systems Based on Behavior". *The SIGART Bulletin*. Vol. 2, No. 4, pp. 46-50.

_____. (1997). "From Earwings to Humans". *Robotics and Autonomous Systems*. Vol. 20, No. 2-4, pp. 291-304.

_____. (1999a). *Cambrian Intelligence. The Early History of the New AI*. Cambridge, MA: The MIT Press.

_____. (1999b). "The Cog Project: Building a Humanoid Robot". In: Nehaniv, C. (ed.) *Computation for Metaphors, Analogy, and Agents*. New York: Springer, pp. 52–87.

_____. (2002). "Humanoid Robots". *Communications of the ACM*. Vol. 45, No. 3, pp. 33-3

_____. (2003). *Flesh and Machines. How Robots Will Change Us*. New York: Vintage.

_____. (2008). "The Next Fifty Years". *Communications of the ACM*. Vol. 51, No. 1, pp. 63-64.

Brooks, R. & Flynn, A. (1989). "Fast, Cheap, and Out of Control. A Robot Invasion of the Solar System." *Journal of the British Interplanetary Society*. Vol. 42, pp. 478-485.

Bruzina, R. (2004). "Phenomenology and Cognitive Science. Moving Beyond the Paradigms." *Husserl Studies*. Vol. 20, No. 1, pp. 43-88.

Buchanan, B. (2005). "A (Very) Brief History of Artificial Intelligence". *AI Magazine*. Vol. 26, No. 4, 53-60.

Campbell-Kelly, M. & Aspray, W. (1996). *Computer. A History of the Information Machine*. New York: Basic Books.

Capurro, R. (2004). "Über Künstlichkeit." In: Kornwachs, K. (ed.) *Technik - System - Verantwortung*. Berlin/Münster/Wien/Zürich/London: LIT, pp. 165-172.

Carter, M. (2007). *Minds and Computers. An Introduction to the Philosophy of Artificial Intelligence*. Edinburgh: Edinburgh University Press.

Carruthers, P. (1996). *Language, Thought and Consciousness. An Essay in Philosophical Psychology*. Cambridge/New York: Cambridge University Press.

Carruthers, P., Stich, S. & Siegal, M. (2002). "Introduction: What Makes Science Possible?" In: Carruthers, P., Stich, S. & Siegal, M. (eds.) *The Cognitive Basis of Science*. Cambridge/New York: Cambridge University Press, pp. 1-19.

Castells, M. (2010a). *The Information Age: Economy, Society, and Culture. Volume I: The Rise of the Network Society*. Malden, MA/Oxford: Wiley-Blackwell.

_____. (2010b). *The Information Age: Economy, Society, and Culture. Volumen II: The Power of Identity*. Malden, MA/Oxford: Wiley-Blackwell.

_____. (2010c). *The Information Age: Economy, Society, Culture. Volume III: End of Millenium*. Malden, MA/Oxford: Wiley-Blackwell.

Chalmers, D. (1996). *The Conscious Mind. In Search of a Fundamental Theory*. Oxford: Oxford University Press.

_____. (1997). "An Exchange with David Chalmers (Reply to John Searle)." In: Searle, J. *The Mystery of Consciousness*. New York: The New York Review of Books, pp. 163-167.

_____. (2000). "Facing up the Problem of Consciousness." In: Schear, J. (ed.) *Explaining Consciousness. The Hard Problem*. Cambridge, MA/London: The MIT Press, pp. 9-30.

_____. (2007). "The Hard Problem of Consciousness." In: Velmans, M. & Schneider, S. (eds.) *The Blackwell Companion to Consciousness*. Malden, MA/Oxford: Blackwell, pp. 225-235.

_____. (2010). "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies*. Vol. 17, No. 9-10, pp. 7-65.

Chapman, D. (1991). *Vision, Instruction, and Action*. Cambridge, MA: The MIT Press.

Chapman, D. et al. (1988). *How to Do Research at the MIT Lab*. AI Working Paper 316. URL: <http://www.cs.indiana.edu/mit.research.how.to.html>. Retrieved: 9/10/2012.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA/London: The MIT Press.

Chorost, M. (2011). *World Wide Mind. The Coming Integration of Humanity, Machines and the Internet*. New York: The Free Press.

Churchland, P. (1981). "Eliminative Materialism and the Propositional Attitudes." In: Rosenthal, D. (ed.) *The Nature of Mind*. New York/Oxford: Oxford University Press. 1991, pp. 601-612.

_____. (1991). "Folk Psychology and the Explanation of Human Behavior." In: Greenwood, J. (ed.) *The Future of Folk Psychology. Intentionality and Cognitive Science*. Cambridge/New York: Cambridge University Press, pp. 51-69.

_____. (1998). "Folk Psychology." In: Churchland, P. & Churchland, P. (eds.) *On the Contrary. Critical Essays 1987-1997*. Cambridge, MA/London: The MIT Press, pp. 3-16.

Clancey, W. (2009). "Scientific Antecedents of Situated Cognition." In: Robbins, P. & Aydede, M. (eds.) *The Cambridge Handbook of Situated Cognition*. Cambridge/New York: Cambridge University Press, pp. 11-34.

Clark, A. (1993). *Microcognition. Philosophy, Cognitive Science, and Parallel Distributed Processing*. Cambridge, MA: The MIT Press.

_____. (1996). "Philosophical Foundations". In: Margaret Boden (ed.) *Artificial Intelligence*. New York: Academic Press, pp. 1-22.

_____. (1997). *Being There. Putting Brain, Body, and World Together Again*. Cambridge, MA/London: The MIT Press.

_____. (2002). "Skills, Spills and the Nature of Mindful Action." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 385-387.

_____. (2012). "Embodied, Embedded, and Extended Cognition." In: Frankish, K. & Ramsey, W. (eds.) *The Cambridge Handbook of Cognitive Science*. Cambridge/New York: Cambridge University Press, pp. 275-291.

Clark, A. & Toribio, J. (2001). "Sensimotor Chauvinism? Commentary on O'Reagan, J. Kevin and Noë, Alva, 'A Sensimotor Account of Vision and Visual Consciousness.'" *Behavioral and Brain Sciences*. Vol. 24, No. 5, pp. 979-980.

Clark, A. & Chalmers, D. (2010). "The Extended Mind." In: Menary, R. (ed.) *The Extended Mind*. Cambridge, MA/London: The MIT Press, pp. 27-42.

Collins, H. (2000). "Four Kinds of Knowledge, Two (or Maybe Three) Kinds of Embodiment, and the Question of Artificial Intelligence." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping and Cognitive*

Science. Essays in Honor of Hubert L. Dreyfus. (Volume 2.) Cambridge, MA/London: The MIT Press, pp. 179-195.
Copeland, J. (1993). *Artificial Intelligence. A Philosophical Introduction*. Oxford: Blackwell.
Costall, A. & Leudar, I. (2009). "Theory of Mind: The Madness in the Method." In: Leudar, I. & Costall, A. (eds.) *Against Theory of Mind*. Basingstoke/New York: Palgrave Macmillan, pp. 39-55.
Crane, T. (2010). *The Mechanical Mind. An Philosophical Introduction to Minds, Machines and Mental Representation*. London/New York: Routledge.
Crevier, D. (1993). *AI. The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books.
Crick, F. (1994). *The Astonishing Hypothesis. The Scientific Search for the Soul*. New York: Touchstone Press.
Cristin, R. (2012). "Phänomenologische Ontologie. Heidegger's Auseinandersetzung mit Husserl (1916-1928)." *Heidegger Jahrbuch*. Vol. 6, pp. 43-68.
Crowell, S. (2012). "Transcendental Phenomenology and the Seductions of Naturalism: Subjectivity, Consciousness, and Meaning." In: Zahavi, D. (ed.) *The Oxford Handbook of Contemporary Phenomenology*. Oxford: Oxford University Press, pp. 25-47.
Cummins, R. & Pollock, J. (eds.) (1991). *Philosophy and AI. Essays at the Interface*. Cambridge, MA: The MIT Press.
Davidson, D. (2001). *Inquiries Into Truth and Interpretation*. Oxford: Clarendon Press.
Deleuze, G. (2006). *Nietzsche and Philosophy*. Trans. by H. Tomlinson. New York: Columbia University Press.
Dennett, D. (1968). "Machine Traces and Protocol Statements." *Behavioral Science*. Vol. 13, No. 2, pp. 155-161.
_____. (1969). *Content and Consciousness*. London: Routledge & Kegan Paul and New York: Humanities Press.
_____. (1971). "Intentional Systems." *Journal of Philosophy*. No. 68, pp. 87-106.
_____. (1978). *Brainstorms*. Cambridge, MA: The MIT Press.
_____. (1982). "How to Study Human Consciousness Empirically or Nothing Comes to Mind." *Synthese*. No. 53, pp.159-180
_____. (1987). *The Intentional Stance*. Cambridge, MA: The MIT Press.
_____. (1988). "When Philosophers Encounter Artificial Intelligence." In: Graubard, S. (ed.) *The Artificial Intelligence Debate. False Starts, Real Foundations*. Cambridge, MA/London: The MIT Press, pp. 283-295.
_____. (1991). *Consciousness Explained*. New York-Boston-London: Back Bay Books.
_____. (1992). "An Interview With Dan Dennett". *Cogito*. Vol. 6, No. 3, pp. 115-125.
_____. (1993a). "Quining Qualia." In: Goldman, A. (ed.) *Readings in the Philosophy of Cognitive Science*. Cambridge, MA · London: The MIT Press, pp. 381-414.
_____. (1993b). "Caveat Emptor: Reply to Three Essays on Consciousness Explained." *Consciousness and Cognition*. Vol. 2, No. 1, pp. 48-57.
_____. (1994a). "The Practical Requirements for Making a Conscious Robot." *Philosophical Transactions: Physical Sciences and Engineering*. Vol. 349, No. 1689, pp. 133-146.
_____. (1994b). "Tiptoeing Past the Covered Wagons: A Response to Carr." *Emory Cognition Project*, report #28, Department of Psychology, Emory University, 2 pp. URL: <http://ase.tufts.edu/cogstud/papers/tiptoe.htm>. Retrieved: 3/24/2011.
_____. (1995). *Darwin's Dangerous Idea. Evolution and the Meanings of Life*. New York: Simon & Schuster.
_____. (1996). "In Defense of AI." Interview with Daniel Dennett. In: Baumgartner, P. & Payr, S. (eds.) *Speaking Minds. Interviews with Twenty Eminent Cognitive Scientists*. Princeton, NJ: Princeton University Press, pp. 59-69.
_____. (1997). "Cog as a Thought Experiment." *Robotics and Autonomous Systems*. Vol. 20, No. 2-4, pp. 251-256.
_____. (1998). *Brainchildren. Essays on Designing Minds*. Cambridge, MA: The MIT University Press.
_____. (2001). "The Fantasy of First-Person Science," written version of a debate with David Chalmers, held at Northwestern University, Evanston, IL,. February 15, 2001, supplemented by an email debate with Alvin Goldman, 11 pp. URL: <http://ase.tufts.edu/cogstud/papers/chalmersdeb3dft.htm>. Retrieved: 3/24/2011.
_____. (2003). "Who's on First? Heterophenomenology Explained." *Journal of Consciousness Studies*. Vol. 10, No. 9, pp. 19-30.

_____. (2005). "Daniel Dennett," self-presentation in Guttenplan, S. (ed.) *A Companion to the Philosophy of Mind*. Massachusetts: Blackwell, pp. 236-244.
_____. (2006). *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA · London: The MIT Press.
_____. (2007). "Heterophenomenology Reconsidered." *Phenomenology and the Cognitive Sciences*. Vol. 6, No. 1-2, pp. 247-270.
_____. (2009a). "Darwin's Strange Inversion of Reasoning." *Proceedings of the National Academy of Science*. Vol. 16, No. 1, pp. 10061-10065.
_____. (2009b). "The Part of Cognitive Science That is Philosophy." *Topics in Cognitive Science*. No. 1, pp. 231-236.
_____. (2012). "A Perfect and Beautiful Machine: What Darwin's Theory of Evolution Reveals about Artificial Intelligence." *The Atlantic*. June 22, 2012. URL: <http://www.theatlantic.com/technology/archive/2012/06/a-perfect-and-beautiful-machine-what-darwins-theory-of-evolution-reveals-about-artificial-intelligence/258829/>. Retrieved: 20/01/2013.
_____. (2013). *Intuition Pumps and Other Tools for Thinking*. New York/London: W. W. Norton & Company.
Dennett, D. & Kinsbourne, M. (1992). "Time and the Observer: the Where and When of Consciousness in the Brain." *Behavioral and Brain Sciences*. Vol. 15, No. 2, pp. 183-247.
Descombes, V. (2001). *The Mind's Provisions. A Critique of Cognitivism*. Trans. by S. A. Schwartz. Princeton/Oxford: Princeton University Press.
Dokic, Jérôme & Elisabeth Pacherie (2007) "Too Much Ado about Belief." *Phenomenology and the Cognitive Sciences*. Vol. 6, No.1-2, pp. 185-200.
Dörner, D. (1999). *Bauplan für eine Seele*. Reinbeck: Rowohlt Verlag.
Dotov D., Nie, L. & Chemero, A. (2010). "A Demonstration of the Transition from Ready-to-Hand to Unready-to-Hand." *Plos One*. Vol. 5, No. 3, pp. 1-9.
Dreyfus, H. (1963). "Wild on Heidegger: Comments." *The Journal of Philosophy*. Vol. 60, No. 22, pp. 677-680.
_____. (1965). *Alchemy and Artificial Intelligence*. RAND Corporation Report P-3244.
_____. (1967). "Why Computers Must Have Bodies in Order to be Intelligent." *Review of Metaphysics*. Vol. 21, No. 1, pp. 13-32.
_____. (1971). "Phenomenology and Mechanism". *Noûs*. Vol. 5, No. 1, pp. 81-96.
_____. (1974). "Artificial Intelligence". *Annals of the American Academy of Political and Social Science*. Vol. 412, pp. 21-33.
_____. (1982). *Husserl, Intentionality, and Cognitive Science*. Cambridge, MA: The MIT Press.
_____. (1987). "From Socrates to Expert Systems: The Limits of Calculative Rationality." *Bulletin of the American Academy of Arts and Sciences*. 40 (4), pp. 15-31.
_____. (1991). *Being-in-the-World. A Commentary on Heidegger's Being and Time, Division I*. Cambridge, MA/London: The MIT Press.
_____. (1992). *What Computers Still Can't Do*. Cambridge, MA: The MIT Press.
_____. (1996a). "Response to my Critics." *Artificial Intelligence*. No. 80, pp. 171-191.
_____. (1996b). "Cognitivism Abandoned." An Interview with Hubert Dreyfus. In: Baumgartner, P. & Payr, S. (eds.) *Speaking Minds. Interviews with Twenty Eminent Cognitive Scientists*. Princeton, NJ: Princeton University Press, pp. 72-83.
_____. (1999). "The Primacy of Phenomenology Over Logical Analysis." *Philosophical Topics*. 27 (2), pp. 3-27.
_____. (2000). "Responses." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping, and Cognitive Science. Essays in Honor of Hubert L. Dreyfus*. (Volume 2.) Cambridge, MA/London: The MIT Press, pp. 313-349.
_____. (2002a). "Intelligence Without Representation: Merleau-Ponty's Critique of Mental Representation. The Relevance of Phenomenology to Scientific Explanation." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 367-383.
_____. (2002b). "Refocusing the Question: Can There Be Skillful Coping Without Propositional Representations or Brain Representations?" *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 413-425.

_____. (2005). "Overcoming the Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise." *Proceedings and Addresses of the American Philosophical Association*. Vol. 79, No. 2, pp. 47-65.
_____. (2007a). "Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian." *Philosophical Psychology*. Vol. 20, No. 2, pp. 247-268.
_____. (2007b) "The Return of the Myth of the Mental." *Inquiry. An Interdisciplinary Journal of Philosophy*. Vol. 50, No. 4, pp. 352-365.
_____. (2007c) "Response to McDowell." *Inquiry. An Interdisciplinary Journal of Philosophy*. Vol. 50, No. 4, pp. 371-377.
_____. (2007d). "Detachment, Involvement, and Rationality: Are We Essentially Rational Animals?" *Human Affairs*. Vol. 17, No. 2, pp. 101-109.
_____. (2012). "A History of First Step Fallacies." *Minds & Machines*. Vol. 22, No. 2, pp. 87-99.
_____. (2013). "The Myth of the Pervasiveness of the Mental." In: Schear, J. (ed.) *Mind, Reason, and Being in the World. The McDowell-Dreyfus Debate*. New York/London: Routledge, pp. 15-40.
Dreyfus, H. & Harrison, H. (eds.) (1982) *Husserl, Intentionality and Cognitive Science*. Cambridge, MA: The MIT Press.
Dreyfus, H. & Dreyfus, S. (1986). *Mind Over Machine. The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
_____. (1988). "Making a Mind vs Modeling the Brain: Artificial Intelligence Back at a Branchpoint." *Dædalus*. Vol. 117, No. 1, pp. 15-43.
Dreyfus, H. & Dennett, D. (1997). "Did Deep Blue's Win Over Kasparov Prove that Artificial Intelligence has Succeeded? A Debate." In: Franchi, S. & Güzeldere, G. (eds.) *Mechanical Bodies, Computational Minds. Artificial Intelligence from Automata to Cyborgs*. Cambridge, MA/London, England: The MIT Press, 2005, pp. 265-279.
Dreyfus, H. & Kelly, S. (2007) "Heterophenomenology: Heavy-Handed Sleight-of-Hand." *Phenomenology and the Cognitive Sciences*. Vol. 6, No. 1-2, pp. 45-55.
Dreyfus, S. (2004) "Totally Model-Free Learned Skillful Coping." *Bulletin of Science, Technology & Society*. Vol. 24, No. 3, pp. 182-187.
Dupuy, J.-P. (2009). *On the Origins of Cognitive Science. The Mechanization of Man*. Trans. by M. B. DeBevoise. Cambridge, MA/London: The MIT Press.
Dukas, R. & Ratcliffe, J. (eds.) (2009). *Cognitive Ecology II*. Chicago/London: The University of Chicago Press.
Dyson, G. (1997) *Darwin Among the Machines. The Evolution of Global Intelligence*. New York: Basic Books.
Ebinger, M. (2012). *Neurophänomenologie: Ein Oxymoron als Lückenfüller. Transformation der Phänomenologie durch Francisco J. Varela—eine Deformation?* Saarbrücken: AV Akademikerverlag.
Edelman, S. (2008). *Computing the Mind. How the Mind Really Works*. Oxford/New York: Oxford University Press.
Ellrich, L. (2003). "Die Computertechnik als Gegenstand philosophischer Reflexion." In: Sandbothe, M. & Nagl, L. (eds.) *Systematische Medienphilosophie*. Berlin: Akademie-Verlag.
Evans G. (1982). *The Varieties of Reference*. Oxford/New York: Clarendon Press and Oxford University Press.
Eysenck, M. (1991). "Consciousness." In: Michael Eysenck (ed.) *The Blackwell Dictionary of Cognitive Psychology*. Oxford/Cambridge, MA: Blackwell, pp. 84-85.
Faingenbaum, G. (2001). *Conversations with John Searle*. Books Online. ISBN 9871022115
Faust, D. & Meehl, P. (2002). "Using Meta-Scientific Studies to Clarify or Resolve Questions in the Philosophy and History of Science." *Philosophy of Science*. Vol. 69, No. 179, pp. 185-196.
Feigenbaum, E. (1977). "The Art of Artificial Intelligence: Themes and Cases Studies of Knowledge Engineering." *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Cambridge, MA, pp. 1014-1029.
_____. (1992). *A Personal View of Expert Systems: Looking Back and Looking Ahead*. Knowledge Systems Laboratory Report No. KSL 92-41. Stanford University.
Feigenbaum, E. & Feldman, J. (1963). *Computers and Thought*. New York: McGraw-Hill.
Feigenbaum, E. & McCorduck, P. (1983). *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*. London: Addison-Wesley.
Fetzer, J. (1990). *Artificial Intelligence: Its Scope and Limits*. Heidelberg: Springer.
Figal, G. (1996). *Der Sinn des Verstehens. Beiträge zur hermeneutischen Philosophie*. Stuttgart: Philipp Reclam.

Flanagan, O. (1991). *The Science of Mind*. Cambridge, MA: The MIT Press.
Flores, F. (1981). *Management and Communication in the Office of the Future*. Phd Dissertation. University of California, Berkeley.
Flores, F. & Ludlow, J. (1981). "Doing and Speaking in the Office." In: Fick, G. & Sprague, R. (eds.) *DSS: Issues and Challenges*. London: Pergamon.
Flores, F. & Graves, M. (1986). "Domains of Permanent Human Concerns and Education." Unpublished reports. Emeryville, CA: Logonet Inc.
Flores, F. & Graves, M. (1988). "Computer Systems and the Design of Organizational Interaction." *ACM Transactions on Office Information Systems*. Vol. 6, No. 2, pp. 153-172.
Floridi, L. (2002). What is the Philosophy of Information? In: Moor, J. & Bynum, T. W. (eds.) *Cyberphilosophy. The Intersection Between Computing and Philosophy*. Malden, MA/Oxford: Blackwell, pp. 117-138.
_____. (ed.) (2008). *Philosophy of Computing and Information. 5 Questions*. Copenhagen: Automatic Press.
_____. (2010). *Information. A Very Short Introduction*. Oxford/New York: Oxford University Press.
_____. (2012). *The Philosophy of Information*. Oxford: Oxford University Press.
_____. (2014). *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*. Oxford/New York: Oxford University Press.
Fodor, J. (1968). "The Appeal to Tacit Knowledge in Psychological Explanation." *The Journal of Philosophy*. Vol. 65, No. 20, pp. 627-640.
_____. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.
_____. (1984). "Semantics, Wisconsin Style." *Synthèse*. Vol. 59, No. 3, pp. 231-250.
_____. (1992). "The Big Idea: Can There Be a Science of Mind?" *Times Literary Supplement*. No. 3, p. 5.
_____. (1996). "The Folly of Simulation." An Interview with Jerry Fodor. In: Peter Baumgartner & Sabine Payr (eds.) *Speaking Minds. Interviews with Twenty Eminent Cognitive Scientists*. Princeton, NJ: Princeton University Press, pp. 85-100.
_____. (2008). *LOT 2. The Language of Thought Revisited*. Oxford: Clarendon Press.
Føllesdal, D. (1979). "Husserl and Heidegger on the Role of Actions in the Constitution of the World." In: Esa Saarinen et al. (eds.) *Essays in Honor of Jaakko Hintikka*. Dordrecht: D. Reidel.
Ford, K., Glymour, C. & Hayes, P. (eds.) (1995). *Android Epistemology*. Cambridge, London: AAAI Press and The MIT Press.
Fox, J. (1996). "Expert Systems and Theories of Knowledge." In: Boden, M. (ed.) *Artificial Intelligence*. New York: Academic Press, pp. 157-181.
Franchi, S. (2006). "Herbert Simon, anti-Philosopher." In: Magnani, L. (ed.) *Computing and Philosophy*. Pavia: Associated International Press, pp. 27-40.
Franchi, S. & Güzeldere, G. (2005). "Machinations of the Mind: Cybernetics and Artificial Intelligence from Automata to Cyborgs." In: Franchi, S. & Güzeldere, G. (eds.) *Mechanical Bodies, Computational Minds. Artificial Intelligence from Automata to Cyborgs*. Cambridge, MA/London: The MIT Press, pp. 15-149.
Freeman, W. (1991). "The Physiology of Perception." *Scientific American*. Vol. 264, No. 2, pp. 78-85.
_____. (1999). *How Brains Make Up Their Minds*. London: Phoenix.
_____. (2000). "A Proposed Name for Aperiotic Brain Activity: Stochastic Chaos." *Neural Networks*. Vol. 13, No. 1, pp. 11-13.
_____. (2008). "Three Types of State Transition Underlying Perception." In: Liljenstrom, H. & Ahrem, P. (eds.) *Consciousness Transitions. Phylogenetic, Ontogenetic and Physiological Aspects*. Amsterdam/Oxford: Elsevier, pp. 231-248.
Frith, C. & Rees, G. (2007). "A Brief History of the Scientific Approach to the Study of Consciousness." In: Velmans, M. & Schneider, S. (eds.) *The Blackwell Companion to Consciousness*. Malden, MA/Oxford: Blackwell, pp. 9-22.
Gadamer, H.-G. (GW 10). *Hermeneutik im Rückblick*. Tübingen: J. C. B. Mohr (Paul Siebeck), 1987.
Gallagher, S. (2003). "Phenomenology and Experimental Design: Toward a Phenomenologically Enlightened Experimental Science." In: Jack, A. & Roepstorff, A. (eds.) (2003) *Trusting the Subject? The Use of Introspective Evidence in Cognitive Science*. (Volume 1.) Exeter: Imprint Academic, pp. 85-99.
_____. (2009). "Philosophical Antecedents of Situated Cognition." In: Philip Robbins & Murat Aydede (eds.) *The Cambridge Handbook of Situated Cognition*. Cambridge/New York: Cambridge University Press, pp. 35-51.

_____. (2010). "Phenomenology and Non-Reductionist Cognitive Science." In: Gallagher, S. & Schmicking, D. (eds.) *Handbook of Phenomenology and Cognitive Science*. New York/Heidelberg/London: Springer, pp. 21-34.
_____. (2012). "On the Possibility of Naturalizing Phenomenology." In: Zahavi, D. (ed.) *The Oxford Handbook of Contemporary Phenomenology*. Oxford: Oxford University Press, pp. 70-93.
_____. (2014). "The Cruel and Unusual Phenomenology of Solitary Confinement." *Frontiers in Psychology*. Vol. 5, No. 12, pp. 1-8.
Gallagher, S. & Varela, F. (2001). "Redrawing the Map and Resetting the Time: Phenomenology and the Cognitive Sciences." In: Crowell, S., Embree, L. & Julian, S. (eds.) *The Reach of Reflection: Issues for Phenomenology's Second Century*. West Harford: Electron Press, pp. 17-43.
Gallagher, S. & Zahavi, D. (2010). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London/New York: Routledge.
Gams, M., Paprzycki, M. & Wu, X. (eds.) (1997). *Mind Versus Computer. Were Dreyfus and Winograd Right?* Amsterdam: IOS Press.
Gardner, H. (1987). *The Mind's New Science. A History of the Cognitive Revolution*. New York: Basic Books.
Gibson, J. (1986). *The Ecological Approach to Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.
_____. (1997). "The Theory of Affordances." In: Shaw, R. & Bransford, J. (eds.) *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum.
Gilson, E. (1999). *The Unity of Philosophical Experience*. San Francisco, CA: Ignatius Press.
Glazebrook, T. (2000). *Heidegger's Philosophy of Science*. New York: Fordham University Press.
Gomi, T. (1997). "Aspects of Non-Cartesian Robotics." *Artificial Life and Robotics*. Vol. 1, No. 2, pp. 95-103.
Goodale, M. et al. (1991). "A Neurological Dissociation Between Perceiving Objects and Grasping Them." *Nature*. Vol. 349, No. 6305, pp. 154-156.
Grondin, J. (2000). "Continental or Hermeneutical Philosophy: The Tragedies of Understanding in the Analytic and Continental Perspectives." In: Sallis, J. & Scott, C. (eds.) *Interrogating the Tradition. Hermeneutics and the History of Philosophy*. New York: SUNY Press, pp. 75-83.
Grush, R. & Mandik, P. (2002) "Representational Parts." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 389-394.
Gurwitsch, A. (1974). *Phenomenology and the Theory of Science*. Ed. L. Embree. Evanston, IL: Northwestern University Press.
_____. (1976). *Studies in Phenomenology and Psychology*. Evanston, IL: Northwestern University Press.
_____. (1979). *Human Encounters in the Social World*. Trans. F. Kernsten. Pittsburgh: Duquesne University Press.
Guzmán, A. (1968). "Decomposition of a Visual Scene into Three-Dimensional Bodies." *Proceedings of the AFIPS Fall Joint Computer Conference*. Vol. 22, pp. 291-304.
Guzzoni, U. (2012). "*Gelassenheit*: Beyond Techno-Scientific Thinking." In: Glazebrook, T. (ed.) *Heidegger on Science*. New York: SUNY Press, pp. 193-204.
Haraway, D. (1988). "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies*. Vol. 14, No. 3, pp. 575-599.
Haugeland, J. (1989). *Artificial Intelligence. The Very Idea*. Cambridge, MA: The MIT Press.
_____. (1996a). "Mind Design." In: Haugeland, J. (ed.) (1997) *Mind Design II. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA/London: The MIT Press, pp. 1-28.
_____. (1996b). "An Overview of the Frame Problem." In: Zenon Pylyshyn (ed.) *The Robot's Dilemma Revisited. The Frame Problem in Artificial Intelligence*. New York: Ablex, pp. 77-93.
_____. (1996c). "Body and World: a Review of *What Computers Still Can't Do. A Critique of Artificial Reason*." *Artificial Intelligence*. No. 80, pp. 119-128.
_____. (1996d). "Farewell to GOFAI?" An Interview with John Haugeland. In: Baumgartner, P. & Payr, S. (eds.) *Speaking Minds. Interviews with Twenty Eminent Cognitive Scientists*. Princeton, NJ: Princeton University Press, pp. 101-114.
_____. (1998). *Having Thought. Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press.
_____. (2013). *Dasein Disclosed*. Ed. by J. Rouse. Cambridge, MA/London: Harvard University Press.
Hawking, S. & Mlodinow, L. (2010). *The Grand Design*. New York: Bantam Books.
Heckmann, H.-D. & Walter, S. (eds.) (2006) *Qualia. Ausgewählte Beiträge*. Münster: Mentis.

Heidegger, M. (GA 1) *Frühe Schriften*. Gesamtausgabe Bd.1. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 1978.

_____. (GA 5). *Holzwege*. Gesamtausgabe Bd. 5. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 2003. [*Off the Beaten Track*. Trans. by J. Young & K. Haynes. Cambridge · New York: Cambridge University Press, 2002.]

_____. (GA 7). *Vorträge und Aufsätze*. Gesamtausgabe Bd. 7. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 2000. [*The Question Concerning Technology and Other Essays*. Trans. by W. Lovitt. New York · London: Garland, 1977].

_____. (GA 8). *Was heisst Denken?* Gesamtausgabe Bd. 8. Ed. by P.-L. Coriando. Frankfurt am Main: Vittorio Klostermann, 2002. [*What is Called Thinking?* Trans. J. G. Gray. New York: Harper & Row, 1968.]

_____. (GA 9). *Wegmarken*. Gesamtausgabe Bd. 9. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 1976.

_____. (GA 14). *Zur Sache des Denkens*. Gesamtausgabe Bd. 14. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 2007. [*On Time and Being*. Trans. by J. Stambaugh. New York: Harper & Row, 1972.]

_____. (GA 16). *Reden und andere Zeugnisse eines Lebensweges*. Ed. by H. Heidegger. Frankfurt am Main: Vittorio Klostermann, 2000.

_____. (GA 17). *Einführung in die phänomenologische Forschung*. [WS 1923-1924]. Gesamtausgabe Bd. 17. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 1994.

_____. (GA 20). *Prolegomena zur Geschichte des Zeitbegriffs*. [SS 1925]. Gesamtausgabe Bd. 20. Ed. by P. Jäger. Frankfurt am Main: Vittorio Klostermann, 1994. [*History of the Concept of Time: Prolegomena*. Trans. by T. Kisiel. Bloomington · Indianapolis: Indiana University Press, 1985.]

_____. (GA 21). *Logik. Die Frage nach der Wahrheit*. [WS 1925-1926]. Gesamtausgabe Bd. 21. Ed. by W. Biemel. Frankfurt am Main: Vittorio Klostermann, 1976.

_____. (GA 24). *Die Grundprobleme der Phänomenologie*. [SS 1927]. Gesamtausgabe Bd. 24. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 1975. [*The Basic Problems of Phenomenology*. Trans. by A. Hofstadter. Bloomington: Indiana University Press, 1982].

_____. (GA 29/30). *Die Grundbegriffe der Metaphysik. Welt - Endlichkeit - Einsamkeit*. [1929-1930]. Gesamtausgabe Bd. 29/30. Ed. by F.-W. von Herrmann. Frankfurt am Main: Vittorio Klostermann, 1983. [*The Fundamental Problems of Metaphysics. World - Finitude - Solitude*. Trans. by W. McNeill & N. Walker. Bloomington · Indianapolis: Indiana University Press, 1995.]

_____. (GA 40). *Einführung in die Metaphysik*. Ed. by P. Jaeger. Frankfurt am Main: Vittorio Klostermann, 1983. [*Introduction to Metaphysics*. Trans. by G. Fried & R. Polt. New Haven · London: Yale University Press, 2000.]

_____. (GA 56/57). *Zur Bestimmung der Philosophie*. [KNS 1919]. Gesamtausgabe Bde. 56/57. Ed. by B. Heimbüchel. Frankfurt am Main: Vittorio Klostermann, 1987.

_____. (GA 58) *Grundprobleme der Phänomenologie*. [WS 1919-1920]. Gesamtausgabe Bd. 58. Ed. by H.-H. Gander. Frankfurt am Main: Vittorio Klostermann, 1993.

_____. (GA 59). *Phänomenologie der Anschauung und des Ausdrucks. Theorie der philosophischen Begriffsbildung*. [SS 1920]. Gesamtausgabe Bd. 59. Ed. by C. Strube. Frankfurt am Main: Vittorio Klostermann, 1993. [*Phenomenology of Intuition and Expression. Theory of Philosophical Concept Formation*. Trans. by T. Colony. New York · London: Continuum, 2010.]

_____. (GA 60). *Phänomenologie des religiösen Lebens*. [WS 1920-1921]. Gesamtausgabe Bd. 60. Ed. by M. Jung, T. Regehly & C. Strube. Frankfurt am Main: Vittorio Klostermann, 1995.

_____. (GA 61). *Phänomenologische Interpretationen zu Aristoteles. Einführung in die phänomenologische Forschung*. [WS 1921-1922]. Gesamtausgabe Bd. 61. Ed. by W. Bröcker & K. Bröcker-Oltmanns. Frankfurt am Main: Vittorio Klostermann, 1994.

_____. (GA 63). *Ontologie. Hermeneutik der Faktizität*. [SS 1923]. Gesamtausgabe Bd. 63. Ed. by K. Bröcker-Oltmanns. Frankfurt am Main: Vittorio Klostermann, 1988.

_____. (SZ). *Sein und Zeit*. [1927]. Tübingen: Max Niemeyer Verlag, 1979. [*Being and Time*. Trans. by J. Macquarrie & E. Robinson. Oxford/Cambridge, MA: Blackwell, 2001.]

_____. (BW). "Die Bedrohung der Wissenschaft." In: Dietrich Papenfuss & Otto Pöggeler (eds.) *Zur philosophischen Aktualität Heideggers (Symposium der Alexander von Humboldt-Stiftung vom 24.-28. April 1989*

in Bonn-Bad Godesberg). Band I. Philosophie und Politik. Frankfurt am Main: Vittorio Klostermann, 1991, pp. 5-27.

_____. (Zoll). *Zollikoner Seminare. Protokolle-Zwiegespräche-Briefe*. Ed. by M. Boss. Frankfurt am Main: Vittorio Klostermann, 1994.

Hendler, J. (2008). "Avoiding Another AI Winter". *IEEE Intelligent Systems*. Vol. 23, No. 2, 1-4.

Herrmann, F.-W. von (1987). *Hermeneutische Phänomenologie des Daseins. Eine Erläuterung von »Sein und Zeit«. Bd. I: Einleitung: Die Exposition der Frage nach dem Sinn von Sein*. Frankfurt am Main: Vittorio Klostermann.

_____. (2000). *Hermeneutik und Reflexion. Der Begriff der Phänomenologie bei Heidegger und Husserl*. Frankfurt am Main: Vittorio Klostermann.

Hobbes, T. (2005). *Leviathan*. (Parts I and II). Ed. A. P. Martinich. Ontario: Broadview Press.

Hume, D. (2000). *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects*. New York · Oxford: Oxford University Press.

Husbands, P., Holland, O. & Wheeler, M. (eds.) (2008) *The Mechanical Mind in History*. Cambridge, MA/ London: The MIT Press.

Husserl, E. (Hua I). *Cartesianische Meditationen und Pariser Vorträge*. Husserliana Bd. I. Ed. by S. Strasser. The Hague: Martinus Nijhoff. 1963. [*Cartesian Meditations*. Trans. by D. Cairns. The Hague · Boston · London: Martinus Nijhoff, 1982.]

_____. (Hua II). *Die Idee der Phänomenologie. Fünf Vorlesungen*. Husserliana Bd. II. Ed. by W. Biemel. The Hague: Martinus Nijhoff. 1973.

_____. (Hua III). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. (Erstes Buch: Allgemeine Einführung in die reine Phänomenologie)*. Husserliana Bd. III. Ed. by W. Biemel. The Hague: Martinus Nijhoff. 1950. [*Ideas. General Introduction to Pure Phenomenology*. Trans. by W. R. Boyce Gibson. London/New York: Routledge, 2002.]

_____. (Hua IV). *Ideen zur einer reinen Phänomenologie und phänomenologischen Philosophie. (Zweites Buch: Phänomenologische Untersuchungen zur Konstitution.)* Husserliana Bd. IV. Ed. by M. Biemel. The Hague: Martinus Nijhoff, 1952.

_____. (Hua V). *Ideen zur einer reinen Phänomenologie und phänomenologischen Philosophie.* (*Drittes Buch: Die Phänomenologie und die Fundamente der Wissenschaften.)* Husserliana Bd. V. Ed. by W. Biemel. The Hague: Martinus Nijhoff, 1971.

_____. (Hua VI). *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie. Eine Einleitung in die phänomenologische Philosophie*. Husserliana Bd. VI. Ed. by W. Biemel. The Hague: Martinus Nijhoff, 1954.

_____. (Hua VII). *Erste Philosophie. Erster Teil: Kritische Ideengeschichte*. Husserliana Bd. VII. Ed. by R. Boehm. The Hague: Martinus Nijhoff, 1956.

_____. (Hua IX). *Phänomenologische Psychologie. Vorlesungen Sommersemester 1925*. Husserliana Bd. IX. Ed. by W. Biemel. The Hague: Martinus Nijhoff, 1968. [*Phenomenological Psychology. Lectures, Summer Semester 1925*. Trans. by J. Scanlon. The Hague: Martinus Nijhoff. 1977.]

_____. (Hua XIX/1). *Logische Untersuchung en. Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*. (Zweiter Band, Erster Teil). Husserliana Bd. XIX/1. Ed. by U. Panzer. The Hague · Boston · Lancaster: Martinus Nijhoff, 1984.

_____. (Hua XIX/2). *Logische Untersuchungen. Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*. (Zweiter Band, Zweiter Teil). Husserliana Bd. XIX/2. Ed. by U. Panzer. The Hague · Boston · Lancaster: Martinus Nijhoff, 1984.

_____. (Hua XXIV). *Einleitung in die Logik und Erkenntnistheorie*. Husserliana Bd. XXIV. Ed. by U. Melle. The Hague: Martinus Nijhoff, 1984.

_____. (Hua XXV). *Vorträge und Aufsätze (1911-1921)*. Husserliana Bd. XXV. Ed. by T. Nenon & H. R. Sepp. Dordrecht · London · Lancaster: Martinus Nijhoff, 1987.

_____. (Hua XXXVII). *Einleitung in die Ethik. Vorlesungen Sommersemester 1920 und 1924*. Husserliana Bd. XXXVII. Ed. by H. Peucker. Dordrecht: Kluwer, 2004.

Hutto, D. & Myin, E. (2013). *Radicalizing Enactivism. Basic Minds Without Content*. Cambridge, MA/ London: The MIT Press.

Jack, A. & Roepstorff, A. (eds.) (2003). *Trusting the Subject? The Use of Introspective Evidence in Cognitive Science*. (Volume 1.) Exeter: Imprint Academic.

_____. (eds.) (2004). *Trusting the Subject? The Use of Introspective Evidence in Cognitive Science*. (Volume 2.) Exeter: Imprint Academic.

Jackson, F. (2002). "Memory Traces and Representation." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 409-410.

Jaén, I. & Simon, J. J. (eds.) (2012). *Cognitive Literary Studies. Current Themes and New Directions*. Austin, TX: University of Texas Press.

Kelly, S. (2000). "Grasping at Straws: Motor Intentionality and Cognitive Science of Skilled Behavior." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping and Cognitive Science. Essays in Honor of Hubert L. Dreyfus*. (Volume 2.) Cambridge, MA · London: The MIT Press, pp. 161-177.

_____. (2001). "Demonstrative Concepts and Experience." *The Philosophical Review*. Vol. 110, No. 3, pp. 397-420.

Kihlstrom, J. (1987). "The Cognitive Unconscious." *Science*. Vol. 237, No. 4821, pp. 1445-1452.

_____. (1999) "Conscious versus Unconscious Cognition." In: Sternberg, R. (ed) *The Nature of Cognition*. Cambridge, MA/London: The MIT Press, pp. 173-203.

Kilian, H. (1970). "Überlegungen zur Metanoetik. Ein Beitrag zur kritischen Theorie unbewusster Strukturen des bewussten Denkens." In: Steinbuch, K. & Moser, S. (eds.) *Philosophie und Kybernetik*. Munich: Nymphenburger Verlangshandlung, pp. 94-121.

Kisiel, T. (2012). "A Supratheoretical Prescientific Hermeneutics of Scientific Discovery." In: Glazerbrook, T. (ed.) *Heidegger on Science*. New York: SUNY Press, pp. 239-259.

Kiverstein, J. (2012). "What is Heideggerian Cognitive Science?" In: Kiverstein, J. & Wheeler, M. (eds.) *Heidegger and Cognitive Science*. Basingstoke/New York: Palgrave Macmillan, pp. 1-61.

Koch Wah, T. (2007). "Heterophenomenology Debunked." *Humanities & Social Sciences*. Vol. I, No. 1, pp. 1-11.

Kuhn, T. (1996). *The Structure of Scientific Revolutions*. Chicago: The Chicago University Press.

Kurzweil, R. (1999). *The Age of Spiritual Machines. When Computers Exceed Human Intelligence*. New York: Penguin.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh. The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Langdridge, D. (2007). *Phenomenological Psychology. Theory, Research and Method*. London/New York: Prentice Hall.

Lask, E. (2003). *Die Logik der Philosophie und die Kategorienlehre / Die Lehre vom Urteil*. Sämtliche Werke Bd. 2. Jena: Dietrich Scheglmann.

Lembeck, K.-H. (1999). "Seinsformen. Spielarten des Ontologiebegriffs in der Phänomenologie Husserls." In: Sepp, H. R. (ed.) *Metamorphose der Phänomenologie. Dreizehn Stadien von Husserl aus*. Freiburg · München: Alber, pp. 28-57.

_____. (2010). *Philosophie als Zumutung? Ihre Rolle im Kanon der Wissenschaften*. Würzburg: Königshausen & Neumann.

Lenat, D., Prakash, M. & Shepherd, M. (1986). "CYC: Using Commonsense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks." *AI Magazine*. Vol. 6, No. 4, pp. 65-85.

Levy, G. (2013). *Judaic Technologies of the Word. A Cognitive Analysis of Jewish Cultural Formation*. Durham: Acumen.

Lewis, C. I. (1929) *Mind and the World. Outline of a Theory of Knowledge*. New York: Charles Scribner's Sons.

Lighthill, J. (1973). "Artificial Intelligence: A General Survey". *Artificial Intelligence: A Paper Symposium*. Science Research Council.

Liu, J. & Newsome, W. T. (2000). "Somatosensation: Touching the Mind's Fingers." *Current Biology*. Vol. 10, No. 16, pp. R598-600.

Livingston, P. (2005). "Functionalism and Logical Analysis." In: Smith, D. W. & Thomasson, A. (eds.) *Phenomenology and the Philosophy of Mind*. Oxford/New York: Clarendon Press/Oxford.

Looren de Jong, H. & Schouten, M. (2005). "Ruthless Reductionism. A Review Essay of John Bickle's Philosophy and Neuroscience: a Ruthless reductive account." *Philosophical Psychology*. Vol. 18, No. 4, pp. 473-486.

Lovejoy, A. (1936). *The Great Chain of Being*. Cambridge, MA: Harvard University Press.

Lungarella, M. et al. (eds.) (2007). *50 Years of Artificial Intelligence. Essays Dedicated to the 50th Anniversary of Artificial Intelligence*. Berlin/Heidelberg/New York: Springer.

Lutterbie, J. (2011). *Toward a General Theory of Acting. Cognitive Science and Performance*. Basingstoke/New York: Palgrave Macmillan.

Lycan, W. (1986). *The Disappearance of Introspection*. Cambridge, MA/London: The MIT Press.

Marbach, E. (1993). *Mental Representation and Consciousness. Towards a Phenomenological Theory of Representation and Reference*. Dordrecht: Kluwer.
_____. (2013). "'Naturalisierung des Geistes' oder 'Natur und Geist'?" *Metodo. International Studies in Phenomenology and Philosophy*. Vol. 1, No. 1, pp. 1-13.
Margolis, E. & Laurence, S. (eds.) (1999). *Concepts: Core Readings*. Cambridge, MA/London: The MIT Press.
Marks, J. (2008). "Great Chain of Being." In: Moore, J. (ed.) *Encyclopedia of Race and Racism*. Michigan: Mcmillan, pp. 68-73.
Marion, J.-L. (1998). *Reduction and Givennes. Investigations of Husserl, Heidegger and Phenomenology*. Trans. by T. A. Carlson. Illinois: Northwestern University Press.
Marr, D. (2010). *Vision. Computational Investigation Into the Human Representation and Processing of Visual Information*. Cambridge, MA/London: The MIT Press.
Marres, R. (1989). *In Defense of Mentalism. A Critical Review of the Philosophy of Mind*. Amsterdam: Rodopi.
Marsh, L. (2007). "Michael Wheeler: Reconstructing the Cognitive World. The Next Step." *Phenomenology and the Cognitive Sciences*. Vol. 7, No. 1, pp. 147-149.
Massey, I. (2009). *The Neural Imagination. Aesthetic and Neuroscientific Approaches to the Arts*. Austin, TX: University of Texas Press.
Matthen, M. (2005). *Seeing, Doing, and Knowing. A Philosophical Theory of Sense Perception*. Oxford/New York: Oxford University Press.
Maturana, H. & Varela, F. (1980). *Autopoiesis and Cognition. The Realization of the Living*. Dordrecht: Reidel.
McCarthy, J. (1955). "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence." *AI Magazine*. Vol. 27, No. 4, 2006, pp. 12-14.
_____. (1958). *Programs with Common Sense*. MIT Laboratory Report Memo 17.
_____. (1996). "Hubert Dreyfus, *What Computers Still Can't Do*." *Artificial Intelligence*. No. 80, pp. 143-150.
McCarthy, J. & Hayes, P. (1969). "Some Philosophical Problems from the Standpoint of Artificial Intelligence." In: *Machine Intelligence 4*. Ed. by D. Michie & B. Meltzer. Edinburgh: Edinburgh University Press, pp. 463-502.
McClamrock, R. (1995). *Existential Cognition. Computational Minds in the World*. Chicago/London: The University of Chicago Press.
McConachie, B. (2008). *Engaging Audiences. A Cognitive Approach to Spectating in the Theatre*. Basingstoke/ New York: Palgrave Macmillan.
McCorduck, P. (1981). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. New York: W. H. Freeman.
McCulloch, W. & Pitts, W. (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity." In: Margaret Boden (ed.) (1990) *The Philosophy of Artificial Intelligence*. Oxford: Oxford University Press, pp. 22-39.
McDermott, D. (1996). "We've been Framed: Or, Why AI is Innocent of the Frame Problem." In: Zenon Pylyshyn (ed.) *The Robot's Dilemma Revisited. The Frame Problem in Artificial Intelligence*. New York: Ablex, pp. 113-122.
McDowell, J. (1994). "The Content of Perceptual Experience." *The Philosophical Quarterly*. Vol. 44, No. 175, pp. 190-205.
_____. (2000). *Mind and World*. Cambridge, MA/London: Harvard University Press.
_____. (2007a). "What Myth?" *Inquiry. An Interdisciplinary Journal of Philosophy*. Vol. 50, No. 4, pp. 338-351.
_____. (2007b). "Response to Dreyfus." *Inquiry. An Interdisciplinary Journal of Philosophy*. Vol. 50, No. 4, pp. 366-370.
_____. (2013). "The Myth of the Mind as Detached." In: Schear, J. (ed.) *Mind, Reason, and Being in the World. The McDowell-Dreyfus Debate*. New York/London: Routledge, pp. 41-58.
McGinn, C. (1991). *The Problem of Consciousness*. Oxford: Blackwell.
McIntyre, R. (1986). "Husserl and the Representational Theory of Mind." *Topoi*. No. 5, pp. 101-113.
McManus, D. (2007). "Rules, Regression and the 'Background:' Dreyfus, Heidegger and McDowell." *European Journal of Philosophy*. Vol. 16, No. 3, pp. 432-458.
Medler, D. (1998). "A Brief History of Connectionism." *Neural Computing Surveys*. Vol. 1, pp. 61-101.
Merleau-Ponty, M. (1996). *The Structure of Behavior*. Trans. by L. A. Fischer. Boston: Beacon Press.
_____. (2005). *Phenomenology of Perception*. New York: Routledge & Kegan Paul.

_____. (2007). "The Primacy of Perception and Its Philosophical Consequences." In: Toadvine, T. & Lawlor, L. (eds.) *The Merleau-Ponty Reader*. Evanston, IL: Northwestern University Press.
Metzinger, T. (2003). *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: The MIT Press.
_____. (2007). "Editorial." *Journal of Consciousness Studies*. Vol. 4, No. pp. 5-6.
Metzinger, T. (ed.) (2006). *Grundkurs Philosophie des Geistes 1: Phenomenales Bewusstsein*. Münster: Mentis.
Michel, J. (2011). *Der qualitativer Charakter bewusster Erlebnisse. Physikalismus und phenomenale Eigenschaften in der analytischen Philosophie des Geistes*. Münster: Mentis.
Mill, J. S. (1994). *The Logic of the Moral Sciences*. Chicago: Open Court.
Millikan, R. G. (2001). *Language, Thought, and Other Biological Categories*. Cambridge, MA/London: The MIT Press.
Minsky, M. (1965). *Matter, Mind and Models*. MIT AI Laboratory Memo 77.
_____. (1967). *Computation. Finite and Infinite Machines*. Eaglewood Cliffs, NJ: Prentice-Hall.
_____. (ed.) (1969). *Semantic Information Processing*. Cambridge, MA: The MIT Press.
_____. (1974). "A Framework for Representing Knowledge." In: Haugeland, J. (ed.) (1997) *Mind Design II. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA/London: The MIT Press, pp. 111-142.
_____. (1988). *Society of Mind*. New York. Simon & Schuster.
_____. (2003). "Why AI is Brain-Dead." Interview with M. Minsky. *Wired*. Vol. 11, No. 8. URL: <http://www.wired.com/wired/archive/11.08/view.html?pg=3>. Retrieved: 15/1/2012.
_____. (2007a). *The Emotion Machine. Commonsense Thinking, Artificial Intelligence and the Future of the Human Mind*. New York: Simon & Schuster.
_____. (2007b). "Discover Interview: Marvin Minsky." Interviewed by Susan Kruglinski. January 2007 Issue. URL: <http://discovermagazine.com/2007/jan/interview-minsky>. Retrieved: 2/6/2014.
Minsky, M. & Papert, S. (1970). *Proposal to ARPA for Research on Artificial Intelligence at MIT, 1970-1971*. MIT AI Laboratory Memo 185.
Minsky, M. & Papert, S. (1971). *Proposal to ARPA for Research on Artificial Intelligence at MIT, 1971-1972*. MIT AI Laboratory Memo 245.
Minsky, M. & Papert, S. (1972). *Artificial Intelligence Progress Report*. MIT AI Laboratory Memo 252.
Minsky, M. & Papert, S. (1988). *Perceptrons. An Introduction to Computational Geometry*. Cambridge, MA: The MIT Press.
Molina, E. (2010). "Kant and the Concept of Life." *The New Centennial Review*. Vol. 10, No. 3, pp. 21-36.
Nagel, T. (1974). "What Is It like to be a Bat?" *The Philosophical Review*. No. 84, pp. 435-450.
Nath, R. (2009). *Philosophy of Artificial Intelligence. A Critique of the Mechanistic Theory of Mind*. Florida: Universal Publishers.
Natorp, P. (2013). *Allgemeine Psychologie nach kritischer Methode. (Erstes Buch: Objekt und Methode der Psychologie.*) Ed. by S. Luft. Darmstadt: Wissenschaftliche Buchgesellschaft.
Nie, L., Dotov, D. & Chemero, A. (2011). "Readiness-to-Hand, Extended Cognition, and Multifractality." In: Carlson, L., Hölscher, C. & Shipley, T. (eds.) *Expanding the Space of Cognitive Science. Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*. Austin: Cognitive Science Society, pp. 1835-1840.
Nilsson, N. (1984). *Shakey the Robot*. Technical Note 323. SRI International.
_____. (2010). *The Quest for Artificial Intelligence. A History of Ideas and Achievements*. Cambridge: Cambridge University Press.
Nisbett, R. & Wilson, T. (1977). "Tell Me More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review*. Vol. 84, No. 3, pp. 231-259.
Newell, A. & Simon, H. (1976). "Computer Science as Empirical Inquiry: Symbols and Search". In: Haugeland, J. (ed.) (1997). *Mind Design II. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA/ London: The MIT Press, pp. 81-110.
Newell, A. (1991.) "Metaphors of Mind, Theories of Mind. Should the Humanities Mind?" In: Sheehan, J. & Sosna, M. (ed.) *The Boundaries of Humanity. Humans, Animals and Machines*. Berkeley/Los Angeles/Oxford: University of California Press, pp. 158-197.
_____. (1994). *Unified Theories of Cognition*. (William James Lectures, 1987). Cambridge, MA/London: Harvard University Press.
Noë, A. (2002). *Out of Our Heads. Why You Are Not Your Brain and Other Lessons from the Biology of Consciousness*. New York: Hill and Wang.
_____. (2006). *Action in Perception*. Cambridge, MA/London: The MIT Press.

_____. (2013). "On Overintellectualizing the Intellect." In: Joseph Schear (ed.) *Mind, Reason, and Being-in-the-World. The McDowell-Dreyfus Debate*. London/New York: Routledge, pp. 178-193.
Olafson, F. (1987). *Heidegger and the Philosophy of Mind*. New Haven/London: Yale University Press.
O'Regan, K. & Noë, A. (2001). "A Sensimotor Account of Vision and Visual Perception." *Behavioral and Brain Sciences*. Vol. 24, No. 5, pp. 939-973.
Papert, S. (1968) *The Artificial Intelligence of Hubert Dreyfus: A Budget of Fallacies*. MIT AI Laboratory Memo 154.
Pauen, M., Schütte, M. & Staudachter, A. eds. (2007). *Begriff, Erklärung, Bewusstsein: Neue Beiträge zum Qualia Problem*. Paderborn: Mentis.
Penrose, R. (1990). *The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics*. Oxford/ New York: Oxford University Press.
_____. (1995). *Shadows of the Mind. A Search for the Missing Science of Consciousness*. Oxford/New York: Oxford University Press.
Petit, J.-L. & Berthoz, A. (2008). *The Physiology and Phenomenology of Action*. Trans. by C. Macann. Oxford/New York: Oxford University Press.
Petitot, J., Varela, F., Pachoud, B. & Roy, J.-M. (eds.) (1999). *Naturalizing Phenomenology*. Stanford: Stanford University Press.
Pfeifer, R. & Bongard, J. (2007). *How the Body Shapes the Way We Think. A New View of Intelligence*. Cambridge, MA · London: The MIT Press.
Phillips, E. (1999). *If It Works It's Not AI: A Commercial Look at Artificial Intelligence Startups*. Master Thesis at MIT.
Pinker, S. (1994). *The Language Instinct. How the Mind Creates Language*. London: Penguin.
Pitt, D. (2000). "Mental Representation". *Stanford Encyclopedia of Philosophy*. URL: <http:// plato.stanford.edu/entries/mental-representation/>. Retrieved: 20/2/2012.
Pöggeler, O. (2000). *Hermeneutik der technischen Welt. Eine Heidegger-Interpretation*. Lüneburg: Unibuch.
Poggio, T. (1981). *Marr's Approach to Vision*. MIT Artificial Intelligence Memo 645.
Polanyi, M. (1962). *Personal Knowledge*. London: Routledge & Kegan Paul.
_____. (1966). "The Logic of Tacit Inference." *Philosophy. The Journal of the Royal Institute of Philosophy*. Vol. 41, No. 155, pp. 1-18.
Pollio, H., Henley, T. & Thompson, C. (1997). *The Phenomenology of Everyday Life*. Cambridge/New York: Cambridge University Press.
Pollock, J. (1989). *How to Build a Person. A Prolegomenon*. Cambridge, MA: The MIT Press.
Popper, K. (2002). *Conjectures and Refutations. The Growth of Scientific Knowledge*. London/New York: Routledge.
Preston, B. (1993). "Heidegger and Artificial Intelligence." *Philosophy and Phenomenological Research*. Vol. LIII, No. 1, pp. 43-69.
Price, H. (2012). "Tim Crane and Huw Price Interview Transcription." URL: <http:// www.phil.cam.ac.uk/news_events/Crane_Price_Interview_Transcription.pdf>. Retrieved: 15/7/2013.
Price, D. & Barrell, J. (2012). *Inner Experience and Neuroscience. Merging Both Perspectives*. Cambridge, MA/ London: The MIT Press.
Protevi, J. (2010). "Adding Deleuze to the Equation." *Phenomenology and the Cognitive Sciences*. Vol. 9, No. 3, pp. 417-436.
Proust, J. (1987). "L'intelligence artificielle comme philosophie." *Le Débat*, No. 47, pp. 88-102.
Putnam, H. (1981). *Reason, Truth and History*. Cambridge, MA: Cambridge University Press.
_____. (1988). "Much Ado About Not Very Much." In: Stephen Graubard (ed.) *The Artificial Intelligence Debate. False Starts, Real Foundations*. Cambridge, MA/London: The MIT Press, pp. 269-281.
_____. (1992). *Renewing Philosophy*. Cambridge, MA: Harvard University Press.
Quine, W. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.
Ratcliffe, M. (2007). "The Problem with the Problem of Consciousness." *Synthesis Philosophica*. Vol. 44, No. 2, pp. 483-494.
_____. (2012). "There Can Be No Cognitive Science of Dasein." In: Kiverstein, J. & Wheeler, M. (eds.) *Heidegger and Cognitive Science*. Basingstoke · New York: Palgrave Macmillan, pp. 135-156.
Rehberg, A. (2012). "Heidegger and Cognitive Science. Aporetic Reflections." In: Kiverstein, J. & Wheeler, M. (eds.) *Heidegger and Cognitive Science*. Basingstoke · New York: Palgrave Macmillan, pp. 157-175.

Rey, G. (2002). "Problems with Dreyfus's Dialectic." *Phenomenology and the Cognitive Sciences*. Vol. 1, No. 4, pp. 403-408.
Richards, J. (ed.) (2002). *Are We Spiritual Machines? Ray Kurzweil vs. the Critics of Strong AI*. Seattle, WA: Discovery Institute.
Richter, E. (2007). "Das Dilemma der modernen Hirnforschung." In: Nielsen, C., Steinmann, M. & Töpfer, F. (eds.) *Das Leib-Seele-Problem und die Phänomenologie*. Würzburg: Königshausen & Neumann, pp. 294-316.
Rinofner-Kreidl, S. (2000). *Edmund Husserl. Zeitlichkeit und Intentionalität*. Freiburg/München: Karl Alber.
Riskin, J. (2007). *Genesis Redux. Essays in the History and Philosophy of Artificial Life*. Chicago: Chicago University Press.
Rorty, R. (1991). "Blunder Around for a While." Vol. 13, No. 22, pp. 3-6.
Rouse, J. (2000). "Coping and its Contrasts." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping and Cognitive Science. Essays in Honor of Hubert L. Dreyfus*. (Volume 2.) Cambridge, MA/London: The MIT Press, pp. 7-28.
_____. (2010). "Heidegger's Philosophy of Science." In: Dreyfus, H. & Wrathall, M. (eds.) *A Companion to Heidegger*. Malden, MA/Oxford: Blackwell, pp. 173-189.
_____. (2013). "What is Conceptually Articulated Understanding?" In: Schear, J. (ed.) *Mind, Reason, and Being-in-the-World. The McDowell-Dreyfus Debate*. London/New York: Routledge, pp. 250-271.
Rowlands, M. (2002). "Two Dogmas of Consciousness." In: Noë, A. (ed.) *Is The Visual World a Grand Illusion?* Exeter: Imprint Academic, pp. 158-180.
_____. (2010a). "Consciousness." In: Gallagher, S. & Schmicking, D. (eds.) *Handbook of Phenomenology and Cognitive Science*. Dordrecht/New York/Heidelberg/London: Springer, pp. 85-97.
_____. (2010b). *The New Science of Mind. From Extended Mind to Embodied Phenomenology*. Cambridge, MA/London: The MIT Press.
Rumelhart, D. (1989). "The Architecture of Mind: A Connectionist Approach." In: Haugeland, J. (ed.) *Mind Design II. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA/London: The MIT Press, pp. 205-232.
Ruse, M. (2012). *The Philosophy of Human Evolution*. Cambridge/New York: Cambridge University Press.
Sáez, L. (2006). "Ontology of Events vs Ontology of Facts: About the Current Fissures Between the Continental and Analytic Traditions." *Journal of the British Society for Phenomenology*. Vol. 37, No. 2, pp. 120-137.
Schank, R. & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding. An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Schear, J. (2013). "Are We Essentially Rational Animals?" In: Schear, J. (ed.) *Mind, Reason, and Being-in-the-World. The McDowell-Dreyfus Debate*. London/New York: Routledge, pp. 285-302.
Scheutz, M. (2002). "Computationalism: The Next Generation". In: Scheutz, M. (ed.) *Computationalism. New Directions*. Cambridge, MA/London: The MIT Press.
Schmidhuber, J. (2007). "2006: Celebrating 75 Years of AI—History and Outlook: The Next 25 Years." In: Lungarella, M. et al. (eds.) *50 Years of Artificial Intelligence. Essays Dedicated to the 50th Anniversary of Artificial Intelligence*. Berlin/Heidelberg/New York: Springer, pp. 29-41.
Schnädelbach, H. (1981). "Morbus hermeneuticus: Thesen über eine philosophische Krankheit." *Zeitschrift für Didaktik der Philosophie*. Vol. 1, pp. 3-6.
Searle J. (1976). "A Classification of Illocutionary Acts." *Language in Society*. Vol. 5, No. 1, pp. 1-23.
_____. (1980). "Minds, Brains, and Programs." *The Behavioral and Brain Sciences*. Vol. 3, No. 3, pp. 417-457.
_____. (1985). *Minds, Brains and Science*. Cambridge, MA. Harvard University Press.
_____. (1991). "Response: The Background of Intentionality and Action." In: Lepore, E. & Van Gulick, R. (eds.) *John Searle and his Critics*. Cambridge, MA/London: Blackwell, pp. 289-299.
_____. (1996). *The Construction of Social Reality*. New York: Penguin.
_____. (1997). *The Mystery of Consciousness*. New York: The New York Review of Books.
_____. (1999). *Speech Acts. An Essay in the Philosophy of Language*. Cambridge/New York: Cambridge University Press.
_____. (2000). "The Limits of Phenomenology." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping, and Cognitive Science. Essays in Honor of Hubert L. Dreyfus*. Vol. 2. Cambridge, MA/London: The MIT Press, pp. 71-92.

_____. (2001a). "Meaning, Mind and Reality." *Revue International de Philosophie*. Vol. 2, No. 217, pp. 173-179.
_____. (2001b). "Neither Phenomenological Description Nor Rational Reconstruction: Reply to Dreyfus." *Revue International de Philosophie*. Vol. 2, No. 217, pp. 277-284.
_____. (2005). "The Phenomenological Illusion." In: Reicher, M. & Marek, J. (eds.) *Experience and Analysis. Proceedings of the 27th International Wittgenstein Symposium*. Vienna, pp. 317-336.
_____. (2007). "Biological Naturalism." In: Velmans, M. & Schneider, S. (eds.) *The Blackwell Companion to Consciousness*. Malden, MA/Oxford: Blackwell, pp. 325-334.
_____. (2008a). *Intentionality. An Essay in the Philosophy of Mind*. Cambridge/New York: Cambridge University Press.
_____. (2008b). *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge/New York: Cambridge University Press.
Seife, C. (2007). *Decoding the Universe. How the New Science of Information is Explaining Everything in the Cosmos, from our Brains to Black Holes*. New York: Penguin.
Sellars, W. (1997). *Empiricism and the Philosophy of Mind*. Cambridge, MA/London: Harvard University Press.
Sengers, P. (1995). "From the Belly of the Devil: Critical Theory in Scientific Practice." *Parallax*. Vol. 2, No. 1, pp. 151-159.
Shanahan, M. (1997). *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*. Cambridge, MA: The MIT Press.
_____. (2004). "The Frame Problem". *Stanford Encyclopedia of Philosophy*. URL: <http://plato.stanford.edu/entries/frame-problem/>. Retrieved: 15/12/2011.
Shanon, C. (1948). "A Mathematical Theory of Communication." *The Bell System Technical Journal*. Vol. 27, pp. 379-423, 623-656.
Shortliffe, E. & Cimino, J. (eds.) (2014). *Biomedical Informatics. Computer Applications in Health Care and Biomedicine*. London/Heidelberg/New York/Dordrecht: Springer.
Siewert, C. (2007). "In Favor of (Plain) Phenomenology." *Phenomenology and the Cognitive Sciences*. Vol. 6, No. 1-2, pp. 201-220.
Simon, H. & Newell, A. (1958). "Heuristic Problem Solving: The Next Advance in Operations Research." *Operations Research*. Vol. 6, No. 1, pp. 1-10.
_____. (1962). "Computer Simulation of Human Thinking and Problem Solving." *Monographs of the Society for Research in Children Development*. Vol. 27, No. 2, pp. 137-150.
Simon, H. (1995). "Technology is not the Problem." In: Baumgarten, P. & Payr, S. (eds.) *Speaking Minds: Interviews with Twenty Eminent Cognitive Scientists*. Princeton, NJ: Princeton University Press.
_____. (1996). *The Sciences of the Artificial*. Cambridge, MA/London: The MIT Press.
Skarda, C. & Freeman, W. (1987). "How Brains Make Chaos in Order to Make Sense of the World." *Behavioral and Brain Sciences*. Vol. 10, No. 2, pp. 161-173.
Smolensky, P. (1989). "Connectionist Modeling: Neural Computation/Mental Connections." In: Haugeland, J. (ed.) *Mind Design II. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA/London: The MIT Press, pp. 233-250.
Smith, B. C. (1996). *On the Origin of Objects*. Cambridge, MA/London: The MIT Press.
Smith, D. W. (2011). "Phenomenology." *Stanford Encyclopedia of Philosophy*, 2008, 24 pp. URL: <http://plato.stanford.edu/entries/phenomenology/>. Retrieved: 24/11/2013.
Soldati, G. (2007). "Subjectivity in Heterophenomenology." *Phenomenology and the Cognitive Sciences*. Vol. 6, No. 1-2, pp. 89-98.
Spinosa, C., Flores, F. & Dreyfus, H. (1997). *Disclosing New Worlds. Entrepreneurship, Democratic Action, and the Cultivation of Solidarity*. Cambridge, MA: The MIT Press.
Sorell, T. (1991). *Scientism. Philosophy and the Infatuation with Science*. London: Routledge.
Steels, L. (2007). "Fifty Years of AI: From Symbols to Embodiment—and Back." In: Lungarella, M. et al. (eds.) *50 Years of Artificial Intelligence. Essays Dedicated to the 50th Anniversary of Artificial Intelligence*. Berlin/Heidelberg/New York: Springer, pp. 18-28.
Stein, E. et al. (2006). "Calculemus! New Research Results and Functional Models Regarding Leibniz Calculating Machine and the Binary Calculating Machine." *Foundations of Civil and and Environmental Engineering*. No. 7, pp. 319-332.
Stenmark, M. (1997). "What is Scientism?" *Religious Studies*. Vol. 33, pp. 15-32.

Steward, J., Gapenne, O. & Di Paolo, E. (eds.) (2010). *Enaction. Towards a New Paradigm for Cognitive Science*. Cambridge, MA/London: The MIT Press.

Stockwell, P. (2005). *Cognitive Poetics. An Introduction*. London/New York: Routledge.

Strawson, P. (1992). *Analysis and Metaphysics. An Introduction to Philosophy*. Oxford: Oxford University Press.

Strom, J. & Darden, L. (1996). "Is Artificial Intelligence a Degenerating Program? A Review of Hubert Dreyfus' *What Computers Still Can't Do*." *Artificial Intelligence*. No. 80, pp. 151-170.

Tatossian, A. (2002). *La phénoménologie des psychoses*. Paris: Vrin.

Thompson, E. (2007). *Mind in Life. Biology, Phenomenology, and the Sciences of the Mind*. Cambridge, MA/ London: Harvard University Press.

Tienson, J. (1988). "An Introduction to Connectionism." *The Southern Journal of Philosophy*. Vol. 26, No. 1, pp. 1-16.

Turing, A. (1950). "Computer Machinery and Intelligence." In: Haugeland, J. (ed.) (1997) *Mind Design II. Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA/London: The MIT Press, pp. 29-56.

_____. (2004). *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life Plus the Secrets of Eni*. Oxford: Oxford University Press.

Tye, M. (1997). *Ten Problems of Consciousness. A Representational Theory of the Phenomenal Mind*. Cambridge, MA/London: The MIT Press.

_____. (2007). "Philosophical Problems of Consciousness." In: Velmans, M. & Schneider, S. (eds.) *The Blackwell Companion to Consciousness*. Malden, MA/Oxford: Blackwell, pp. 23-35.

_____. (2011). "Representationalist Theories of Consciousness." In: McLaughlin, B., Beckermann, A. & Walter, S. (eds.) *The Oxford Handbook of Philosophy of Mind*. Oxford: Clarendon Press, pp. 253-267.

Uexküll, J. (2010). *A Foray Into the Worlds of Animals and Humans With a Theory of Meaning*. Trans. by J. O'Neill. Minneapolis/London: University of Minnesota Press.

Varela, F., Thompson, E. & Rosch, E. (1993). *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge, MA/London: The MIT Press.

Varela, F. (1996). "Neurophenomenology. A Methodological Remedy for the Hard Problem." *Journal of Consciousness Studies*. Vol. 3, No. 4, pp. 330-349.

Vedral, V. (2012). *Decoding Reality. The Universe as Quantum Information*. Oxford/New York: Oxford University Press.

Waltz, D. (1972). *Generating Semantic Descriptions from Drawings of Scenes with Shadows*. MIT Laboratory Report 271.

Warrington E. K. & Weiskrantz, L. (1968). "New Method of Testing Long-Term Retention with Special Reference to Amnestic Patients." *Nature*. Vol. 217, No. 132, pp. 972-974.

Weiskrantz, L. & Warrington, E. K. (1975). "Blindsight. Residual Vision Following Occipital Lesions in Man and Monkey." *Brain Research*. No. 85, pp. 184-185.

Welton, D. (2000). *The Other Husserl. The Horizons of Transcendental Phenomenology*. Bloomington: Indiana University Press.

Wheeler, J. A. (1992). "Recent Thinking about the Nature of the Physical World: It from Bit." *Annals of the New York Academy of Sciences*. Vol. 655, pp. 349-364.

Wheeler, M. (1994). "From Activation to Activity: Representation, Computation, and the Dynamics of Neural Network Control Systems." *Artificial Intelligence and Simulation of Behaviour Quaterly*. No. 87, pp. 36-42.

_____. (1995). "Escaping from the Cartesian Mind-Set: Heidegger and Artificial Life." *Lecture Notes in Computer Science*. Vol. 929, pp. 65-76.

_____. (1996). *The Philosophy of Situated Activity*. University of Sussex PhD Thesis.

_____. (2005). *Reconstructing the Cognitive World: The Next Step*. Cambridge: The MIT Press.

_____. (2008). "God's Machines: Descartes on the Mechanization of Mind." In: Husbands, P., Holland, O. & Wheeler, M. (eds.) *The Mechanical Mind in History*. Cambridge, MA/London: The MIT Press, pp. 307-330.

_____. (2012). "Naturalizing Dasein and Other (Alleged) Heresies." In: Kiverstein, J. & Wheeler, M. (eds.) *Heidegger and Cognitive Science*. Basingstoke/New York: Palgrave Macmillan, pp. 176-212.

Whiteside, J. & Wixon, D. (1988). "Contextualism as a Worldview for the Reformation of Meetings." *Proceedings of the Second Conference on Computer-Supported Cooperative Work*. Association for Computing Machinery, pp. 369-376.

Wilson, E. (1978). *On Human Nature*. Cambridge, MA: Cambridge University Press.

Windelband, W. (1909). *Die Philosophie im deutschen Geistesleben des XIX. Jahrhunderts. Fünf Vorlesungen*. Tübingen: J. C. B. Mohr (Paul Siebeck).
Winograd, T. (1972). "Understanding Natural Language." *Cognitive Psychology*. No. 3, pp. 1-191.
_____. (1991). "Thinking Machines. Can There Be? Are We?". In: Sheehan, J. & Sosna, M. (ed.) *The Boundaries of Humanity. Humans, Animals, Machines*. Berkeley/Oxford: University of California Press, pp. 198-223.
_____. (1995). "Heidegger and The Design of Computer Systems." In: Feenberg, A. & Hannay, A. (ed.) *Technology and the Politics of Knowledge*. Bloomington: Indiana University Press, pp. 108-127.
_____. (1996). *Bringing Design to Software*. New York: ACM Press.
Winograd, T. & Flores, F. (1987). *Understanding Computers and Cognition. A New Foundation for Design*. Indianapolis: Addison-Wesley Professional.
Winston, P. (1974). *New Progress in Artificial Intelligence*. MIT AI Laboratory Report 310.
_____. (1975). *Learning Structural Descriptions from Examples*. MIT Laboratory Report 231.
_____. (1992). *Artificial Intelligence*. New York: Addison-Wesley.
Wittgenstein, L. (2000). *Über Gewissheit*. Trans. by J. L. Prades & V. Raga. Barcelona: Gedisa.
Wixon, D. & Wilson, C. (1997). "The Usability Engineering Framework for Product Design and Evaluation." In: Helander, M., Landauer, T. & Prabhu, P. (eds.) *Handbook of Human-Computer Interaction*. Amsterdam/New York: Elsevier, pp. 653-688.
Wolf, G. (2008). "Futurist Ray Kurzweil Pulls Out All the Stops (and Pills) to Live to Witness the Singularity." *Wired Magazine*. Vol. 16, No. 4. URL: <http://www.wired.com/medtech/drugs/magazine/16-04/ff_kurzweil>. Retrieved: 12/12/2012.
Wrathall, M. (2000). "Background Practices, Capacities, and Heideggerian Disclosure." In: Wrathall, M. & Malpas, J. (eds.) *Heidegger, Coping and Cognitive Science. Essays in Honor of Hubert L. Dreyfus*. (Volume 2.) Cambridge, MA/London: The MIT Press, pp. 93-114.
Wright, E. (2003). "Dennett as Illusionist." *Journal of Theoretical and Philosophical Psychology*. Vol. 23, No. 2, pp. 157-167.
_____. (2011). "Why Transparency is Unethical." In: Wright, E. (ed.) *The Case for Qualia*. Cambridge, MA/London: The MIT Press, pp. 341-366.
Zahavi, D. (2003). *Husserl's Phenomenology*. Stanford, CA: Stanford University Press.
_____. (2004). "Husserl's Noema and the Internalism-Externalism Debate." *Inquiry* 47, pp. 42-66.
_____. (2005). "Being Someone." *Psyche*. Vol. 11, No. 5, pp. 1-20.
_____. (2006). "Does (Husserlian) Phenomenology Have a Future?" Internet project organized by the *Husserl Archives* at the *New School for Social Research*. URL: <http://www.newschool.edu/nssr/husserl/Future/Part%20Two/Zahavi.html> Retrieved: 12/12/2013.
_____. (2007). "Killing the Straw Man: Dennett and Phenomenology." *Phenomenology and the Cognitive Sciences*. Vol. 6, No. 1-2, pp. 21-43.
_____. (2008). *Subjectivity and Selfhood. Investigating the First-Person Perspective*. Cambridge, MA/London: The MIT Press, 2008.
_____. (2010). "Naturalized Phenomenology." In: Gallagher, S. & Schmicking, D. (eds.) *Handbook of Phenomenology and Cognitive Science*. New York/Heidelberg/London: Springer, pp. 3-19.
_____. (2013). "Mindedness, Mindfulness, and First-Person Authority." In: Schear, J. (ed.) *Mind, Reason, and Being-in-the-World. The McDowell-Dreyfus Debate*. London/New York: Routledge, pp. 320-343.
Zerubavel, E. (1999). *Social Mindscapes. An Invitation to Cognitive Sociology*. Cambridge, MA: Harvard University Press.

—

Lebenslauf

· Seit 2015. Professor am Institut für Philosophie an der Universität Costa Rica.
· 2015. Promotion zum Dr. phil. an der Julius-Maximilians-Universität Würzburg. Dissertation: *The Primacy of Phenomenology Over Cognitivism. Towards a Critique of the Computational Theory of Mind.*
· 2010-2015. Doktorand an der Graduiertenschule für die Geisteswissenschaften der Julius-Maximilians-Universität Würzburg.
· 2007-2010. Dozent am Institut für Philosophie der Universidad de Costa Rica.
· 2010. Erlernen der deutschen Sprache am Goethe-Institut Freiburg.
· 2007. Lizenziat der Philosophie. Abschlussarbeit zum Thema „Der Begriff der Phänomenologie beim jungen Heidegger."
· 2005. Bachelor der Philosophie.
· 2000-2007. Studium der Philosophie, Latein und Altgriechisch an der Universidad de Costa Rica.
· 2006. Erlernen der deutschen Sprache am Goethe-Institut Freiburg.
· 2002-2005. Erlernen der deutschen Sprache am Goethe-Institut in San José, Costa Rica.
· 1997-2000. Studium der Evangelischen Theologie an der Universidad Bíblica Latinoamericana.
· 1995. Abitur an der zweisprachigen (Spanisch-Englisch) Schule Mont Berkeley International.
· 1977. Geboren in San José, Costa Rica.