Published in: AI & Society Journal of Journal of Knowledge, Culture and Communication 2017, online first

»Super-intelligent« Machine: Technological Exuberance or the Road to Subjection

Prof. Dr.-Ing. Peter Brödner
Baumeisterstr. 48
76137 Karlsruhe
Peter.Broedner@t-online.de

Received: 24 March 2017

Abstract

Looking back on the development of computer technology, particularly in the context of manufacturing, we can distinguish three big waves of technological exuberance with a wave length of roughly 30 years: In the first wave, during the 1950es, mainframe computers at that time were conceptualized as »electronic brains« and envisaged as central control unit of an »automatic factory« (Wiener). Thirty years later, during the 1980es, knowledge-based systems in computer-integrated manufacturing (CIM) were adored as the computational core of the »unmanned factory«. Both waves dismally stranded on the contumacies of reality. Nevertheless, again thirty years later, we now experience the departure of the »smart factory« based on networks of »artificially intelligent« multi-agent or »cyber-physical systems« (often also addressed as »internet of things«).
From the very beginning, these technological exuberances rooted in mistaken metaphors describing the artifacts (e.g. »electronic brain«, »knowledge-based« or »intelligent systems«) and, hence, in delusions about the true nature of computer systems. The behaviour of computers is, as computing science teaches us, strictly restrained to executing computable functions by means of algorithms, it thus neither resembles the performance of a brain as part of a complex sensitive living body nor is it in any meaningful sense »knowledgeable« or »intelligent« (this predicate remaining reserved for the programmer designing the algorithms). When the delusion of being able to implement »smart factories«, despite the countless accomplishment failures before, gains momentum anew, it appears absolutely essential to reflect on underlying misconceptions.

Key words:

Automatic factory, cyber-physical systems, multi-agent systems, artificial neural networks, big data, functionalism, praxeological perspective.

1 Introduction: Dreaming of the Automatic Factory

With the ubiquitous proclamation of »industry 4.0« and »big data« as core of a new »industrial revolution«, we experience another wave of technological exuberance propagating advanced computer technology as a panacea for all sorts of societal problems from poor resource efficiency to demographic imbalances. Hardly any trouble appears big enough not to be overcome by »digitization«. Against the background of knowledge-intensive value creation, continuously weak growth and decreasing productivity growth, the World Economic Forum reminds in a report, to undertake major efforts specifically for gaining higher competitiveness in manufacturing and services by improving performance and control over globally dispersed value adding chains by means of the »digital transformation« – seemingly the original document of the movement (WEF 2012).

In accord with these widely shared ideas, highly developed and industrialized countries adopted more or less effective measures for their advancement; particularly, Germany with its strong industry base launched a framework programme for the advancement of »innovation in manufacturing and services« as central part of their »high tech« development strategy. Production components such as »intelligent« machines, work pieces, and storage systems are envisaged to form globally networked »multi-agent« or »cyber-physical systems« (CPS). Enabled by advanced computer technology, these systems can automatically exchange data and mutually initiate actions between the components in »decentralized self-organisation« and, thus, accomplish »smart factories and services«, sufficiently adaptive and dynamic for economically producing individual customer orders, handling disturbances and failures, and optimal decision making (BMBF 2014).

More generally, in their new book Brynjolfsson and McAfee (2014) describe the advent of a »second machine age« and how the new digital revolution changes the world. Referring to the extraordinary exponential growth of computer performance and storage capacity in digital networks (according to Moore's law), with rapid progress in accomplishing »artificial intelligence« and »big data« applications, they illuminate the potential for »digital« value creation. For explanation, they refer to Google's the self-driving car and IBM's Watson with respect to knowledge processing capacities. In critical perspective, they also address, however, the risks of monopolizing digital value creation due to network effects (»the winner takes it all«), look at expanding inequalities and polarities of skills and income, and discuss opportunities for controlling future developments.

This in many respects evokes memories of earlier attempts to make come true management's old dream of an automatic factory, of eternal unmanned »value creation« by means of computer and sensor technology, the dream of finally becoming independent of the obstinacy and contumacy of living labour. As early as 1950, when the first commercial mainframe computers had just been installed, Norbert Wiener (1950) already had a clear and detailed vision how to achieve an automatic factory by means of sensors, effectors and computing machines as central logical units for controlling its complex processes. And thirty years later, during the 1980es, the central idea of the »unmanned factory« directed high development efforts into »knowledge-based« (i.e. equipped with symbolic »artificial intelligence«) and computer-integrated manufacturing (CIM) systems (Brödner 1990, Hunt 1989). These tidal waves of technological exuberance, arriving with a length of roughly 30 years, each time dismally stranded at the cliffs of unruly matter and underestimated implementation problems with ensuing long phases of disillusionment when trying to overcome those difficulties. Ironically, in each of these sobering phases, the value of implicit knowledge, of intuition and creativity, of specifically human acting skills was rediscovered.

The paper wants to realistically assess the similarities and differences of the new promises of the »second machine age«, particularly »industry 4.0« and »big data«, relative to previous attempts. To this end, the paper starts with presenting the scientific and technical foundations of the various attempts to accomplish an automatic factory until present in some more detail. By comparing the attempts, the novelty of the most recent approach can be determined. Based on that, a critical evaluation of the opportunity and risk potentials can be made. Finally, realistic design perspectives for forward looking manufacturing and service systems will be derived.

2 Revenant Symptoms: The Third Wave

2.1 Previous attempts of creating an »unmanned factory«

Looking back on the development of computer technology, particularly in the context of manufacturing, we can distinguish three big waves of technological exuberance with a wave length of roughly 30 years: In the first wave, during the 1950es, mainframe computers at that time were conceptualized as »electronic brains« and envisaged as central control unit of an »automatic factory«:

> »The computing machine represents the center of the factory, but it will never be the whole factory. On one hand, it receives its detailed instructions from elements of the nature of sense organs. […] Besides these sense organs, the control system must contain effectors or components which act on the outer world.
> Of course, we assume that the instruments which act as sense organs record not only the original state of the work, but also the result of all previous processes. Thus the machine may carry out feedback operations, either those of the simple type now so thoroughly understood, or those involving more complicated processes of discrimination, regulated by the central control as a logical or mathematical system. In other words, the all-over system will correspond to the complete animal with sense organs, effectors, and proprioceptors, and not, as in the ultra-rapid computing machine, to an isolated brain, dependent for its experiences and for its effectiveness on our intervention« (Wiener 1950, 156f).

Despite the persuasive power of Wiener's clearly outlined vision of the automatic factory, computers did not penetrate manufacturing to a considerable degree until the mid 1970es. There were of course, some cases of early investment in, for instance, numerically controlled machine tools and computerized management of a firm's material (and money) flows as well as R&D activities in computer-aided design (all being areas with already highly standardized operations and procedures). However, difficulties in getting access to isolated mainframe computers prevented wide-spread use. With the advent of »virtual machine« operating systems (IBM 360/370, DEC PDP 10/11) providing computer access via locally dispersed terminals, the use of computing power in manufacturing gained considerable momentum. Far from implementing fully automated operations , however, computing machinery was rather used in a more or less interactive mode combining computing functions with skilled human expert work.

Meanwhile, during the high time of Taylorism with its separation of planning from operating, huge amounts of explicit propositional knowledge about optimal operating conditions and procedures in manufacturing had been collected. Additionally, it turned out that, with rapidly expanding computer programs in many manufacturing areas (NC machines, computer-aided production planning and control, computer-aided design, cost accounting), many of these programs used the same data. In order to avoid error-prone multiple data entries, the idea arose to integrate the many programs and software components so far used in isolation into »computer-integrated manufacturing« systems (CIM) by means of a common data base. Moreover, as at the same time market forces and competition changed their nature from standardized mass to flexible quality production, and against the background of a wealth of explicit manufacturing knowledge at hand, the integration idea was combined with efforts to develop symbolic artificial intelligence using this knowledge for automatically handling the complex and dynamic, steadily changing operating procedures. The intention was to widely replace skilled shop-floor and knowledge workers by »knowledge-based systems«. »Experts leave, while expert systems remain« was a common slogan at the time. This is how thirty years after Wiener's vision, during the 1980es, »knowledge-based« and »expert systems« in computer-integrated manufacturing were promoted as guiding ideas and computational core of the newly envisaged »unmanned factory« (for more details cf. Brödner 1990, 2007, Hunt 1989).

Both waves were fueled by a technology-centred perspective which culpably ignored essential conditions for successful performance of manufacturing processes. In particular, it disregarded deep societal changes such as the transition from industrial to knowledge-based economies and, hence, the relevance of social relationships and the division of labour and knowledge for efficient value creation. As knowledge work is becoming a dominant factor in manufacturing and services (Bell 1973, Drucker 1994), it is important to get a clear understanding of the fundamental differences between implicit practical competence and explicit conceptual or propositional knowledge and how they interplay with each other (for basic differences cf. Nonaka 1996, Polanyi 1966, Ryle 1949). While the individually embodied action competence as pre-reflexive implicit knowledge or working capacity is always antecedent and expresses itself in activities of successful social practice, explicit propositional knowledge *about* certain aspects of this practice can only be gained through observation, concept formation and analysis. However, this codified knowledge needs to be made effective again by appropriating it for practical use (this also holds for technical artifacts derived from this knowledge). Both, explicating propositional knowledge about competences and appropriating it for practical use, are skillful activities expressing the working capacity which for his part is augmented through these activities. Ironically and contrary to common expectations, these dynamic relationships have just precisely been illuminated by the difficulties of »knowledge elicitation« for building expert systems (Brödner 2013).

Consequently, both waves dismally stranded on the contumacies of reality. In fact, the efforts to build knowledge-based integrated manufacturing systems ended in a complete reversal: Confronted with examples of real high performance manufacturing systems, it became obvious that these systems, contrary to common belief, produced their high efficiency not primarily by means of advanced computer technology, but rather through skillful cooperation between human experts in multifunctional teams such as cellular group work or simultaneous engineering teams. Exactly this could also be concluded from theoretical insight in the dynamic relationship between the – always partial – explication of practical experience and action competence into codified knowledge and the appropriation of this knowledge for practical problem solving. The latter typically requires expertise and knowledge from diverse domains to be consolidated and integrated through the self-organized cooperation of experts. The more differentiated, complex, and dynamic the codified knowledge is – and its objectification in technical artifacts –, the more demanding competence and working capacity are required to seize hold of these productive forces for effective practical use. This is subject to the experts' autonomy and cannot be planned and instructed.

Both waves were also accompanied by apocalyptical predictions of lasting technological unemployment. As the shop-floor worker before, now also the knowledge worker would be replaced to a substantial degree by computing machinery. Electronic data processing was generally denounced as severe »job killer«. Nothing of this really happened, though; instead, a productivity paradox could be observed with computer use: »You can see the the computer age everywhere, but in productivity statistics« (Solow 1987). Although the rapidly expanding use of computers in manufacturing and services, growing both in volume and diversity of applications, caused massive changes in professions, specific skills, and qualifications, the macroeconomic productivity effect was minimal. In fact, we experience, since a number of years, a secular downturn in macroeconomic productivity growth rates (Gordon 2014). On firm level, however, huge differences in performance between firms operating under comparable market conditions and with similar software applications in use could be observed. Two decades of empirical research efforts investigating these effects finally came, in accord with theoretical view above, to the conclusion:

> »To leverage information technology investments successfully, firms must typically make large complementary investments and innovations in areas such as business organization, workplace practices, human capital, and intangible capital.« (Jorgenson et al. 2008, 10; similarly also Dedrick et al. 2003).

2.2 The new machinists' claims

Notwithstanding these rather sobering experiences, we now see, again thirty years later, the departure of a »smart factory« based on networks of »artificially intelligent« multi-agent or »cyber-physical systems« (CPS; often also addressed as »internet of things«). With the striking designations »industry 4.0« or »second machine age«, respectively, this is intended to mark another qualitative leap in industrial development. With its focus on advances in computer technology, it again indicates a new wave of technocentrism and technological exuberance. So it is worth while to throw a closer look on the scientific and technological foundations.

Embedded systems are computer components for digital control of physical processes which are equipped with interfaces to humans and other components. By data exchange via the internet, they can be globally networked (cyber-physical systems«, »internet of things and services« (cf. fig. 1).

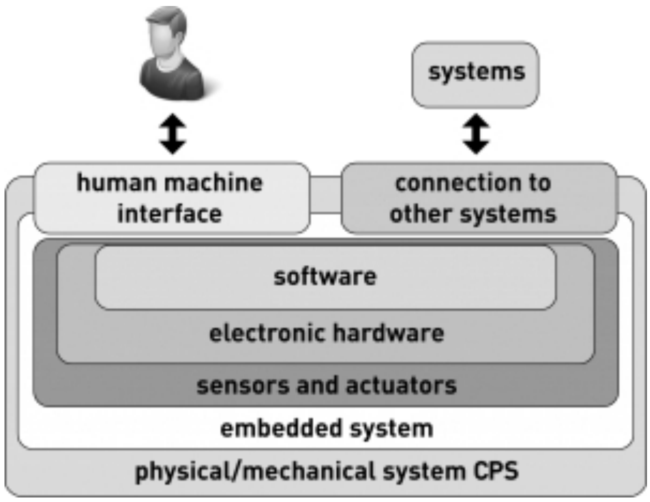

Fig. 1: Embedded system (according to Broy)

Multi-agent systems (MAS, also called »distributed artificial intelligence«) consist of software-agents with limited autonomy for goal-oriented interaction by data transfer; such concerted action enables them to jointly master demanding tasks.

For a deeper understanding, a closer look on the behaviour of these systems, their mode of operation and the way they are described is required. The extraordinary interest in MAS goes back to the idea that coordinated action of a large number of units with relative simple behaviour would produce »artificial intelligence« (Minsky 1988). It is based on the explicitly articulated conviction that »interaction is more powerful than algorithms« (Wegner 1997). This has, however, immediately been proven wrong, as MAS also underly the constraints of computability (Prasse & Rittgen 1998). As a matter of fact, however, the behaviour of MAS as wholes does show emergent properties that cannot be observed with any single software-agent.

Agents are software engineering objects capable of taking in sensor data from the environment as well as from other agents, of independently processing the data by means of their own algorithms – mostly »machine learning« algorithms –, and of putting out resulting data. The behaviour is characterized by the capacity to follow goals and to adapt to changing conditions by »machine learning« (Breadshaw 1997, Maes 1994). In order to cope with more demanding tasks, the agents with limited autonomy each can thus cooperate for achieving the tasks by concerted action (Wooldridge 2002). Simultaneously, huge amounts of data are being generated which can be used separately.

Although each agent by itself performs relatively simple algorithms and shows transparent behaviour, the MAS as a whole owns a highly complex behaviour which cannot be analysed and understood from outside any more, although it still is strictly determined by algorithms. Formally, CPS and MAS can be described as so called »non-trivial machines« (Foerster 1991) whose output data are not only determined by its input data, but also by its variable internal state that is itself a function of input data. The internal state reflects the various ways in which the agents interact and adapt their behaviour (cf. fig. 2). Consequently, the MAS behaviour highly depends on the history and cannot be analytically determined from outside and, hence, foreseen any more.

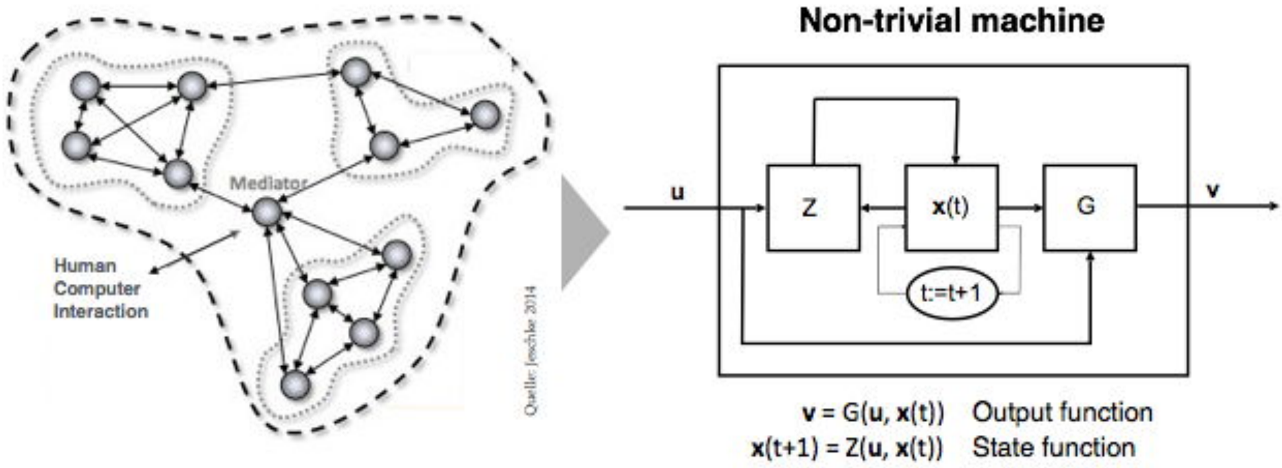

Fig. 2: MAS as non-trivial machine

Besides this physical and algorithmic description of the MAS behaviour, another type of description is also commonly used which orients itself at the agents' purposive interactions according to the so called »intentional stance« (Dennett 1987). In this stance, intentional states like convictions, beliefs, desires or intentions are ascribed to agents expressively in order to simplify the description of the agents' behaviour. This common practice among MAS researchers refers to a report by McCarthy (1979):

> »To ascribe beliefs, free will, intentions, consciousness, abilities, or wants to a machine is legitimate when such an ascription expresses the same information about the machine that it expresses about a person. It is useful when the ascription helps us understand the structure of the machine, its past or future behaviour, or how to repair or improve it. It is perhaps never logically required even for humans, but expressing reasonably briefly what is actually known about the state of the machine in a particular situation may require mental qualities or qualities isomorphic to them. Theories of belief, knowledge and wanting can be constructed for machines in a simpler setting than for humans, and later applied to humans. Ascription of mental qualities is most straightforward for machines of known structure such as thermostats and computer operating systems, but is most useful when applied to entities whose structure is incompletely known.« (McCarthy 1979, quoted according to Shoham 1993, 53).

An agent's intentions are formally described by means of propositional modal logic which augments the normal two-value logic (a proposition can be either true or false) by the two modalities that a proposition is necessarily or only possibly true or false. On this augmented logical basis, an agent's intentional states like beliefs, desires or intentions can then be expressed in a formal language. The fact e.g. that agent a has the conviction that proposition ϕ is true at time t can thus be modelled by the formal language expression $\mathbf{Bel}(a,\phi)(t)$ (Wooldridge 2002).

The abstraction from real physical and algorithmic behaviour by ascribing intentionality to software agents or machines is an attempt to escape from the unpleasant fact that the course of this behaviour is untransparent, although determined. Because the behaviour cannot be explained on the physical level, it is pretended, by abstraction, to underly intentionality. This lastly absurd kind of quasi-explanation leads to believe only that comprehension is possible, while the real behaviour still defies understanding. Rather than analyzing the problem, it is obscured instead. Both, however, the missing behavioural transparency as well as the attempt of ascribing intentionality as quasi-explanation, have fatal consequences for human-machine interaction with and safety of MAS (Norman 1994).

The vision of the »smart factory« highly depends on »machine learning« capabilities for adaptive behavior. They comprise diverse methods and algorithmic procedures for purposefully changing an agent's structure, or its program respectively, such that its behavior is improved relative to a given utility function. The present chief attraction of the »smart factory« and »smart service« propagandists is »deep learning« in so-called »artificial neural networks« (ANN; for an introduction cf. Kriesel 2007). They consist of many simple, connected processors called neurons, each producing a sequence of real-valued activations according to specific computing functions resembling neural functions (cf. fig. 3).

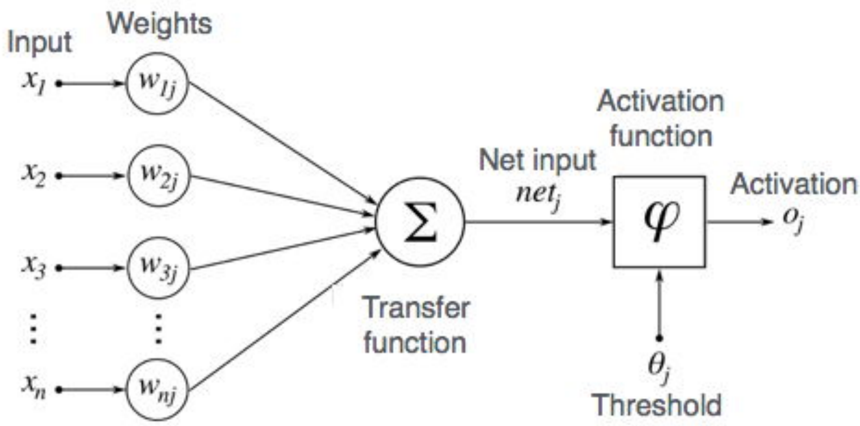

Fig. 3: Computing functions of a«neural« processor forming a node

Input neurons get activated through sensors perceiving the environment, while other neurons get activated in a layered order through weighted connections from previously active neurons. Some neurons may influence the environment by triggering actions. »Learning« then is about finding weights that make the neural network exhibit desired behaviour. Depending on the problem and how the neurons are connected, such behaviour may require long sequences of algorithmically controlled computational stages, where each stage transforms the aggregate activation of the network. For problems of speech recognition or image categorization e.g. – forming tasks where ANN are specifically successful – a very long sequence of patterns or images is presented as input together with the correct categories at the output; from this assignment, the network can, in many small adaptive steps, automatically compute the connecting weights w_{ij} (which implicitly mirror the learning progress) (cf. fig. 4).

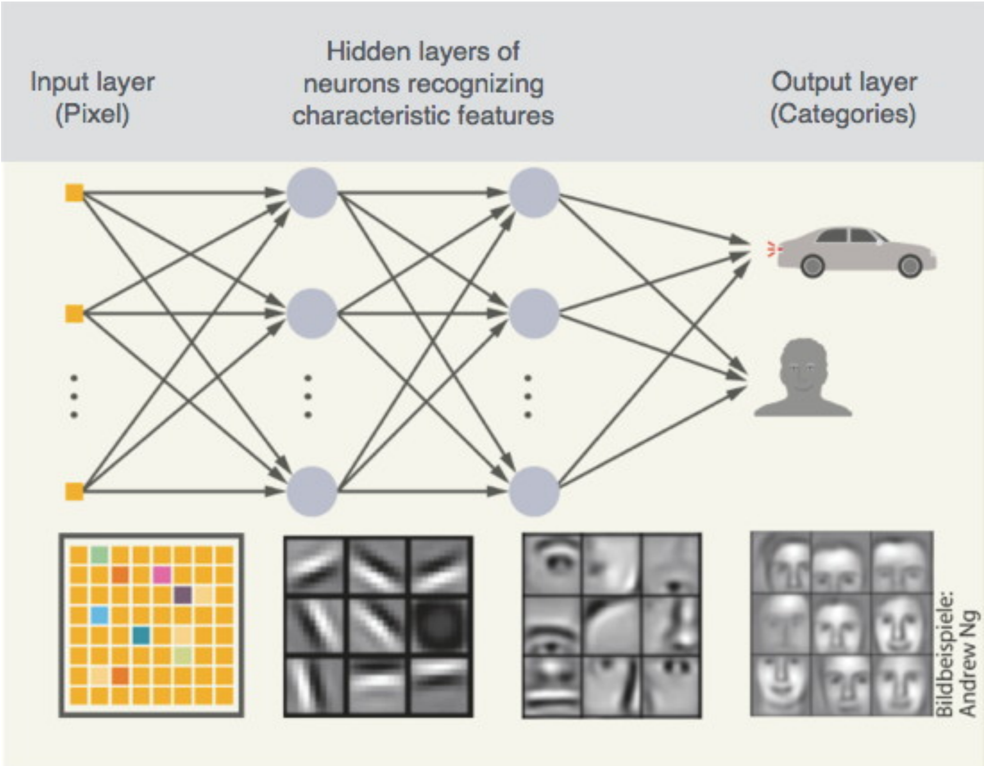

Fig. 4: ANN for categorizing images (Source: c't 6/2016)

Consequently, successful ANN deployment does not depend on analytical insight into cause and effect relationships but rather is the result of theory-less trial and error with different structures and learning algorithms for adjusting the weights. Due to their extraordinary performance, so called »convolutional neural networks« whose structures are inspired by biological models have moved into the focus of interest for the time being.

Contrary to what is suggested, ANN are not conceptually new computational devices at all – in fact, they first appeared in 1950es and have been developed and proven since (with a long »winter« of disinterest in-between). Progress is not accomplished by any conceptually new AI idea but predominantly by the exponentially increased computational power allowing for much larger networks with considerably more components and layers (obviously in the illusion that quantity somehow turns into quality). Ironically, good results are mainly achieved, however, by the intuition, skills, and experience of the developers in structuring the networks and mastering the numerous practical computational problems – from the »vanishing gradient« to the variable increment control of learning algorithms (Schmidhuber 2015).

In practical operation, users have to cope with the uncertainty whether the computed results are correct and suitable in the long run. Even if the ANN deliver adequate results in the vast majority of cases, they may suddenly fail without notice by the users. Even slightly disturbed input data can lead to considerable failures (Sharif et al. 2016). For lack of transparency of the non-linear behavior of ANN, its reliability is difficult to evaluate; basically, users have no chance but to blindly trust in their functionality.

Moreover, the appropriation for interactive use is seriously troubled in case of multi-agent systems with »deep learning« facility. Formally, these systems are »non-trivial machines« whose behaviour depends on history and, therefore, is intransparent and unforeseeable, although algorithmically determined. How can human actors put such systems to deliberate instrumental use that each time exhibit a different behaviour – a property that clearly contradicts the HCI requirement of expectation conformity? On the users' side, excessive expectations for the systems' alleged »action competence« would be evoked at the same time. Confronted with this kind of contradictions, simultaneously exposed to high pressure of management's expectations for successfully mastering their tasks, despite the loss of control over their means of work with intransparent behaviour, the workers would suffer from permanent psychic stress (as already analyzed by Norman 1994).

Moreover, the division of tasks and functions between remaining workers and automatically operating computing components is of fundamental significance for sociotechnical systems design. With respect to this essential design aspect, a number of »ironies of automation« have been early revealed from analyzing working activities in central control rooms, which even gain relevance since with growing systems complexity: Automatically operating »learning« systems like MAS are designed to widely replace human expert knowledge workers, whose working capacity is, however, urgently needed in cases of disturbance or failure. The specific skills of the working capacity are fading away, though, to the extent they are not used during normal automatic operations. In the long run, a severe loss of competence will occur which will turn originally highly competent users into helpless, unpracticed »operators« (Bainbridge 1983, for impressive new examples cf. Carr 2013).

Consequently, these systems are inappropriate for interactive use; they can only be designed and operated as self-contained automata with the incalculable risk of undesired behaviour (examples of such »normal accidents« (Perrow 1984) are numerous).

2.3 Persistent self-deception

In order to catch the original speech flair of the »smart factory« propagandists, it is worth while to start with quoting a typical representative, a top manager from a German automotive company, who at a recent conference characterized the specific features of the »Robotic Enterprise: The Future AI Company« in the year 2025 with the following statements:

> »Super intelligent, continuously learning computers will take over much of what humans deal with so far: They automatically respond to customer or supplier questions by means of so called bots, they autonomously decide how prices for more than 200 models are adapted from country to country, they even design cars and compute how they can be produced. They will also take over cost accounting and controlling. […] They will even manage business meetings« (SZ 11.11.2016, own translation).

Comparing this with a similar statement from the previous wave of technological exuberance (or again with Wiener's earlier vision of 1950 quoted above) reveals high accordance:

> »Computer integration represents the core of the future manufacturing innovation. It aims at automatically producing variable production programmes. … A new manufacturing structure emerges which, as a mechanic organism with programmed and, hence, stored intelligence is capable of automatically producing goods. … On this higher development level, the factory will need machine intelligence« (G. Spur 1984, a leading German manufacturing researcher at the time; own translation).

An unmistakable red thread of obviously contrafactual and wishful thinking winds through these visions. From the very beginning, these illusions and technological exuberances rooted in mistaken metaphors describing the envisaged computer artifacts, e.g. »electronic brain«, »artificial intelligence«, »knowledge-based systems«, or »machine learning«. With their analogies and references to specifically human capabilities, the metaphors obscure essential differences between artificially created machines and autonomously living, socially interactive organisms. Hence, they produce delusions about the true nature of computer systems. The behaviour of computers is, as computing science teaches us, strictly restrained to executing computable functions by means of algorithms, it thus neither resembles the performance of a brain as part of a complex sensitive living body nor is it in any meaningful sense »knowledgeable« or »intelligent« – this predicate remaining reserved for the programmer designing the algorithms or the users making sense of the computing functions. When the delusion of being able to implement »smart factories«, despite the countless accomplishment failures before, gains momentum anew, it appears absolutely essential to reflect on underlying misconceptions.

According to Hofstadter & Sander (2013), analogies are at the core of cognition; analogies allow to understand encountered new phenomena by means of existing experiences, they are instruments by which we apply the wealth of our previous experiences to the presence, and without them we would helplessly navigate in the world. Therefore, it is of essential importance to draw on appropriate analogies transferring the predominant characteristic to the new phenomenon. Exactly this fails with the above analogies taking specific human capacities for machine functions, thus confusing the true nature of both. This fallacy ultimately leads to a mistaken equating of both phenomena.

This can be exemplified by ascribing intentionality to machines according to the »intentional stance« (as quoted above). It is legitimated »when such an ascription expresses the same information about the machine that it expresses about a person. It is useful when the ascription helps us understand the structure of the machine, its past or future behaviour, or how to repair or improve it.« The key word here is »information« which itself is totally confusing, as it denominates different, incompatible concepts: either the syntactical measure of the »entropy« of a string of signs from a finite set (alphabet) according to Shannon (1948) or »any difference that makes a difference« in the context of a social practice according to Bateson (1980). By leaving this open, the physical world of deliberately designed machines with prescribed behaviour is confused with the social world of autonomous actors with the faculty of speech, of creating knowledge, and of designing purposeful artifacts.

Similarly, the term »machine learning« is again based on a mistaken analogy or attribution. The machine's changing behaviour is achieved by algorithmic procedures controlling its adaptation to environmental changes (in fact, this type of machines have formerly been rightly called »adaptive systems«). In contrast, human learning is essentially based on reflective action control and the capacity of concept formation as foundation for creating explicit propositional knowledge. This confusion nourishes the illusion that computers equipped with »artificial intelligence« and »deep learning« capacities can widely replace human skills and working capacity.

In philosophical terms, these mistaken metaphors are built on the fluctuant grounds of *functionalism*. Elaborated as an approach to overcome the flaws of behaviourism, functionalism (Putnam 1960, Fodor 1968) recognises mental states as essential internal entities for explaining behaviour. Irrespective of their material implementation (as electronic hardware or a biological brain), the mental states are regarded, however, as purely functional states according to the Turing machine model. This philosophical view is meanwhile resumed to be refuted, though, because equal mental functions can – as Putnam, one of its originators later has shown (1991) with sharp-witted thought experiments – produce totally different real world references like thoughts or experiences. In its more recent variation, with an interpretation of »embodiment« (cf. e.g. Varela et al. 1991) that again is narrowed by a positivistic attitude, CPS are simply equated with sensitive living human bodies which are up to empathy and able to reflect their context-bound experiences (cf. e.g. most recently Jeschke 2015). The scientifically founded difference between algorithmically determined behaviour of an artifact and intentionally controlled sense making action and understanding in the context of social practices is again ignored.

With respect to the use of computers in organizations with knowledge work or value creation, where human experts interact with advanced computer systems, this difference is highly relevant for assessing the machinists' claims of being able to implement »smart factories or services« replacing human competence and expertise. At the core, they deny the fundamental ontological difference between physical events and social facts: While causal relationships in the physical world – in which, on the basis of semiconductor physics and formal logic, machine computation is operating – exist independently of human activity, objects and facts of the social world such as signification, meaning, or institutions are solely created and maintained through communication and cooperation based on shared collective intentionality: They are originated by declaration, i.e. by speech acts that make something the case just by representing it as being the case (Searle 2010).

Consequently, there seemingly is a unsurmountable gap between both worlds with respect to the social construction of meaning being inaccessible for computing machinery unless it is embedded in and appropriated by actors in the social world of an organization. For bridging the gap, it is useful to pick up the triadic sign concept elaborated by C.S. Peirce (1903), an American logician rooted in the pragmatist strand of research (who also was the first to develop a first order predicate calculus as another foundation for computing science). With this sign concept, he distinguishes a physical entity, a signifying element called representamen, from a second entity as the object it refers to or designates, which can be present, distant, or imagined only. A third entity, the interpretant, assigns meaning to this reference in some context of social interaction. This sign concept provides the missing link to connect the physical world, in which computers process signals or data, with the social world of signification, of constructing meaning through interpretation in an action context. Distinct signification or interpretation of the data are possible due to the actors' functional knowledge of the algorithms and of the inputs. Specifically, it opens up a *praxeological perspective* (Reckwitz 2002) for analyzing the complex interplay of algorithmically determined physical data processing with the social process of signifying or interpreting the data in the context of an organization's social practices (this can be done e.g. leaning on Giddens' theory of structuration (1984), for more details cf. Brödner 2009). In this way, the triadic sign concept mediates between signal and sense.

Confusing both worlds – as it is indicated for instance, when meaningless data are constantly equated with meaningful information – has effects in two directions: In one direction, it reduces living beings, in a positivistic and reductionistic attitude, to the functionality of machines, while on the side of the users it creates the illusion of capabilities comparable to theirs. Leaning on this, the »smart factory« approach can, of course, be misused as ideological offensive: With threatening scenarios of replacement, living labour is set under pressure to accept the new »industrial revolution« in all aspects as an inescapable »natural« event. Public awareness thus is distracted from deliberate massive institutional, particularly labour market deregulation which often are the real threat to decent work and income (which is e.g. mostly the case with businesses operating »crowdsourcing« and »clowdworking« platforms). In this way, detrimental consequences can easily be ascribed to »technological progress« which allegedly cannot be »hold up«.

3 Big Data and the Struggle for Autonomy

The pervasive implementation of »cyber-physical systems« is accompanied by origination of huge amounts of data – »big data« – being processed in unprecedented volume, variety, and velocity. These data can originate from diverse sources and they may differ in the way they are structured, e.g. as text or image documents, or data base entries, and can still be combined for processing. Because of the exponential increase in computer power with respect to processor and storage capacity, it is now possible to keep huge amounts of data ready for very fast processing in the random access memory (so called »in-memory technology«). Moreover, highly expanded band widths allow for transferring huge data volumes. If necessary and with suitable tasks, data volumes and algorithms can even be split to different, locally dispersed processors. All this contributes to exploiting the full performance potential of advanced computer technology for coping with complex tasks.

Big data processing gives rise to some substantial, eventually even unsolvable problems, though. One severe problem concerns the methodology of rational processing itself. Only recently, the editor of the internet magazine »Wired« has, in typical, commonly shared attitude of technological exuberance, proclaimed the »end of theory« in a full-bodied way: Theoretically informed research with proven scientific methods could just be replaced by huge volumes of data, in the »petabyte age« forecasts on the basis of pure correlations would be superior to hypotheses-based propositions, and correlation would replace causality (Anderson 2008). This unbelievable folly reproduces the well-known fallacy of *»cum hoc ergo propter hoc«*: When two events *a* and *b* coincide, one can never know without additional expensive analysis whether *a* has been caused by *b* or reversely *b* by *a*; one cannot even know whether both events depend on an unrecognized common third incident or whether they occur just by accident. Ultimately, this view leads to the apophenic delusion of perceiving patterns in purely random data. It looks like the inmates are running the asylum.

Big data volumes, particularly if they stem from different sources, normally have deficient quality: The data mostly are not representative or error-prone, they can even be obsolete or inconsistent. In many cases, one cannot even assess the extent to which the quality is deficient. As long as the big data processing does not apply accepted strong methods for statistical conclusion, however, which include knowledge of the data quality, it must be seen as scientifically embellished reading tea leaves. Nevertheless, data have frequently been adjudged to be an »important economic good«, the »bulk oil of the 21st century«; if so, they then need equal careful and expensive refinery for extracting useful information.

Finally the deficient data security produces severe problems. Organizations run into huge risks by loss or theft of data, by either spying from outside or sabotaging from inside (risks about which almost daily reports on »cyber attacks give evidence). The risks become even higher, if the data and the processing procedures use to be outsourced to service providers or into the »cloud«. With respect to frequency and volume of the damages experienced, it is hard to understand why firms with ambitions for »industry 4.0« projects deliberately expose their continuously emerging, highly competition-sensitive data streams about products and processes to such risks. Even if much technical and organizational efforts is invested in data security, they will never be sufficient, since any sophisticated security measure can, as experience teaches, be overcome again.

According to a bon mot of the grand semiotician Umberto Eco (1976), a sign is »anything that can be used for lying«. This illuminates in a paradoxical way again the deep insight Peirce had into the logic of signs according to which data represent something in a certain context and for a social actor only. Their meaning always is the result of interpretation in the shared context of a social practice. That which a physical sign stands for, the designated object, and the meaning which is ascribed to it, are first of all up to the sign's author using it for a message. Those who take it up for interpretation are free, to interpret it as expected or other (as far as the context allows). In other words: How data are being interpreted defies the author's control. This is why signs can always be used to deceive, to trick, to defame, or to degrade (all this being frequent practices in the social web and by secret services as well). Via the social web, thus otherwise locally constrained practices of intrigue are becoming a global phenomenon.

Physical computer signals or data are, due to their formal and abstract nature, context-free and meaningless; nonetheless, they frequently use to be equated with meaningful information derived from context-dependent interpretation. By this common error it is suggested that physical signals or data as such allegedly possess information, meaning, and validity quasi as fixed qualities. It is true that parts of the context can be reconstructed from a number of data referring to the same object or person if the according algorithm's semantics is known and, thus, constrain the range of interpretation, but this incompletely reconstructed context still leaves space for various other interpretations and misinterpretations. Despite this interpretive space and often questionable data quality, data suggest objectivity and factuality like in cases of presumptive evidence (rightly seen as questionable). Moreover, due to the social construction of reality, the interpretation often describes a reality just created by the signifying process itself or it even unfolds the effect of a social norm: Descriptive can be turned into normative data, frequency can change to certainty, and interest-bound signification can be enforced by power (as can be seen e.g. in cases of self-tracking or of determining creditworthiness; cf. Boyd 2011).

Due to these peculiarities of the social use of signs, control over the data and their processing algorithms delivers a powerful ruling instrument in the hands of management or government. With comprehensive global data collections and various data processing methods, a powerful instrument for behavioural control and dominance emerges in the hands of the owners; it can, at any time, be used in many ways for exerting influence and power at their discretion, from manipulating public opinion to threat and blackmail. This is possible exactly because the suppliers of the data lose the sovereignty and control over their interpretation in the moment they give the data away. This enables the successful implementation of a perfect panopticon (*sensu* Bentham and Foucault) lighting all corners of knowledge work processes or the social web: It remains the secret of the observer whose behaviour he observes and how he is interpreting it.

Besides the severe use problems with non-trivial machines, this loss of informational autonomy appears as the most threatening societal damage the third wave is about to produce. As it might end in a digital totalitarianism, the struggle for autonomy on all levels of social practice is of foremost importance.

4 Conclusion: Perspectives for Intelligence Amplification

Computers are data processing machines, hence their functionality is semiotic in nature. They fundamentally differ from classical machines transforming matter or energy: While the latter operate in the physical world of natural processes and their functionality makes use of natural forces and effects for increased efficiency and productivity, computers perform computable functions within formalised sign structures, processing signals or data determined by algorithms, nothing else. Formalizing sign processes, their reduction to computable functions, therefore is a necessary prerequisite.

When operating in organisations, computers and their »auto-operational forms« (Floyd 2002) are, based on sufficiently modelling and formalizing underlying sign processes, fully embedded in the social world of social interaction in the organization's practice, namely the expanding knowledge work. Computers can, thus, be used to organize, process, and store codified knowledge represented in data (Brödner 2009). Productivity, therefore, can only improve, if the sign processes of this social practice are organized more efficiently through computer operations – this being the true reason for the empirical findings on the productivity paradox quoted above. Unfortunately, these essential relationships are obscured by the unreasonable terms »digitization« and »digital transformation« permanently used for computerizing knowledge work.

Only by emphasizing the fundamental differences, it is possible to adequately design decent and efficient computer-supported work. Recognizing the differences and taking the praxeological perspective as outlined allows to focus the view on how exactly computer artifacts are emerging from conceptually analyzing social practices, how appropriating their functions for effective use intervenes in social practices of knowledge work, and what the decisive issues of taking influence are. Both, the design, predominantly the analysis, modelling, and formalization of sign processes, as well as the organization of the elaborate appropriation of the computer functions derived from that for effective practical use, are the neuralgic fields of participatory intervention of computer experts and potential users. Both are highly contested terrains with respect to interpretation, interests and exertion of power. That is why they also are the main fields of influence of knowledge workers and their stakeholders. Due to its utmost importance as productive force, the development of working capacity of living labour, its implicit knowledge, practical skills, and competences, must be the guiding principle.

In design of sociotechnical systems, in particular in development, implementation, and use of computers in manufacturing, activities have, as historical retrospective shows, always been underlying two contrarian perspectives:

- The *technology-centred perspective* of most extensively automating knowledge work as it is driven by the efforts for accomplishing »artificial intelligence« – *AI (artificial intelligence)* –: »Smart machines« and »autonomous agents« networked as multi-agent systems with »deep learning« capacity and combined with »big data« procedures are envisaged to imitate and widely replace human working capacity in manufacturing and services; their capacity to »learn« – in fact their adaptivity to environmental conditions only – is nevertheless supposed to provide sufficient flexibility for adapting to changing requirements (according to the »intentional stance«; Minsky 1988, Shoham 1993, Wooldridge 2002).
- Under the *praxeological perspective*, in contrast, advanced computer systems, designed, appropriated, and used as human and task appropriate tools and media for cooperation – *IA (intelligence amplification)* –, are envisaged to support living labour in such a way that the working capacity and, consequently, productive and innovative capacities are enabled and stimulated to grow: »things that make us smart« (Norman 1993; cf. also Ehn 1988, Winograd 1996).

As indicated above, due to poor previous experiences and the problems presented with various *AI* efforts, the technology-centred perspective appears to be less promising, rather a waste of resources. In contrast, evidence-based examination reveals that the widespread use and secular success of computer technology is predominantly based on the praxeological *IA* perspective of *intelligence amplification* and the organizational development efforts connected to that approach. In this perspective, human skills, particularly reflective and conceptual learning capacities, are combined with the precision and velocity of the machine. This must be put in the center of awareness in order to combine flexibility with efficiency. The socio-technical design then needs to be oriented at the peculiarities and needs of human acting and social practice. In particular, those conditions need to be regarded, under which human working capacity can unfold for increased productivity and creativity. For accomplishment, working tasks sustaining competence and fostering learning, task-appropriate, transparent and controllable means of work with expected behaviour, as well as sufficient time resources for appropriating the tools and optimizing processes are needed (Brödner 2013). Four decades of extensive labour research provide a sound footing for that (although this knowledge base seems to presently fade away).

According to the dynamics of explicating practical skills as explicit codified knowledge and of appropriating this knowledge as augmented skill, the use of computer systems in organizations massively intervenes into their social practices, frequently with surprising results. Practical human skills and experiences supported by task-appropriate tools often prove to be superior to »smart« automatically operating systems replacing human actors; this is even true, if the automata perform better than human experts: The chess champion Kasparov e.g., who had been outperformed by IBM's »Deep Blue« computer, has on his part beaten again a comparably powerful computer by using a much simpler personal computer as a supportive tool (Kasparov 2010).

According to this type of human-computer interaction, continuously collected data from CPS might, for instance, be used as input in systems for interactive assistance with advanced usability to reconfigure or optimize production processes, to simulate and control such processes, or to use data analytics for preventive maintenance. For effective interaction, it is important that users have opportunity to control the degree of detail with which they can look at the progress of machine or process states, at given settings, or at methods in use. This is needed in order to enable the users to generate a discrete picture of the incidents or the constitution of results, and to purposefully interfere. On the other hand, for handling the data safely, general agreements need to be accepted that regulate their use practice, particularly access conditions and operating methods.

To follow this praxeologically informed IA-perspective, means to accomplish higher flexibility, productivity, and innovation capacity by sociotechnical design of decent work, rather than betting on questionable AI-promises. It means to organize a productive, creative and autonomous cooperation of competent and knowledgeable experts supported by useful and usable computer artifacts such that their working capacity and competence can further grow. It lastly means to leave the road to subjection.

References

Anderson, C. 2008: The End of Theory, Wired 23.06.08

Bainbridge, L. 1983: Ironies of Automation, Automatica 19, 775-779

Bateson, G. (1980) Mind and Nature. A Necessary Unity. Toronto: Bantam Books

Bell, D. 1973: The Coming of the Post-Industrial Society. A Venture in Social Forecasting, New York: Basic Books

BMBF 2014: Innovationen für die Produktion, Dienstleistung und Arbeit von morgen, Bonn (Federal frame programme fostering innovation for manufacturing, services and work of tomorrow)

Boyd, D. 2011: Six Provocations for Big Data, www.softwarestudies.com/cultural_analytics/ Six_Provocations_for_Big_Data.pdf [last access: 10.02.2015]

Breadshaw, J.M. 1997: An Introduction to Software Agents. In: Breadshaw, J.M. & Hutchinson, F. (eds.): Software Agents, Cambridge (MA): MIT Press, 3-46

Brödner, P. 2013: Reflective Design of Technology for Human Needs, 25th Anniversary Volume: A Faustian Exchange: What Is to Be Human in the Era of Ubiquitous Technology, AI & Society Journal of Human-Centred Systems 28, 27-37

Brödner, P. 2009: The Misery of Digital Organisations and the Semiotic Nature of IT, AI & Society Journal of Human-Centred Systems 23, 331-351

Brödner, P. 2007: From Taylorism to Competence-based Production, AI & Society Journal of Human-Centered Systems 21 (4), 497-514

Brödner , P. 1990: The Shape of Future Technology. The Anthropocentric Alternative, London: Springer

Brynjolfsson, E. & McAfee, A. 2014: The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies, New York London: Norton & Comp.

Carr, N. 2013: All Can Be Lost: The Risk of Putting Our Knowledge in the Hands of Machines, The Atlantic No. 11

Dedrick, J.; Gurbaxani, V. & Kraemer, K.L. 2003: Information Technology and Economic Performance: A Critical Review of the Empirical Evidence, ACM Computing Surveys 35, 1-28

Dennett, D.C. 1987: The Intentional Stance, Cambridge (MA): MIT Press

Drucker, P.F. 1994: The Age of Social Transformation, The Atlantic No. 11, 53-80

Eco, U. 1976: A Theory of Semiotics, Bloomington: Indiana University Press

Ehn, P., 1988: Work-Oriented Design of Computer Artifacts, Stockholm: Arbetslivscentrum

Floyd, C., 2002: Developing and Embedding Auto-operational Form, in: Dittrich, Y.; Floyd, C.; Klischewski, R. (Eds.): Social Thinking – Software Practice, Cambridge (MA): MIT Press, 5-28

Fodor, J. 1968: Psychological Explanation, New York: Random House

Foerster, H.v. 1991: Through the Eyes of the Other, in: Steyer, F. (Ed.): Research and Reflexivity, London: Sage Publications, 21-28

Giddens, A. 1984: The Constitution of Society, Outline of the Theory of Structuration, Cambridge: Polity Press

Gordon R.J. 2014: The Demise of US.Economic Growth: Restatement, Rebuttal, and Reflections, NBER Working Paper 19895

Hofstadter D. & Sander, E. 2013: Surfaces and Essences: Analogy as the Fuel and Fire of Thinking, New York: Basic Books

Hunt, V.D. 1989: Computer Integrated Manufacturing Handbook. London New York: Kluwer Academic Publishers

Jeschke, S. 2015: Auf dem Weg zu einer »neuen KI«: Verteilte intelligente Systeme, Informatik Spektrum 38 (1), S. 4-9 (Towards a »new AI«: Distributed intelligent systems)

Jorgenson, D.W.; Ho, M.S. & Stiroh, K.J. 2008: A Retrospective Look at the U.S. Productivity Growth Resurgence, Journal of Economic Perspectives 22 (1), 3-24

Kasparov, G. 2010: The Chess Master and the Computer, The New York Review of Books 11.02.2010

Kriesel, D. 2007: A Brief Introduction to Neural Networks, http://www.dkriesel.com/science/neural_networks [last access: 12.02.2017]

Maes, P., 1994: Agents that Reduce Work and Information Overload, CACM 37 (7), 31-41

Minsky, M. 1988: The Society of Mind, New York: Simon & Schuster

Nonaka, I., 1996: A Dynamic Theory of Organizational Knowledge Creation, Organization Science 5 (1), 14-37

Norman, D.A. 1994: How Might People Interact with Agents, CACM 37 (7), 68-71

Norman, D.A. 1993: Things that Make Us Smart, Reading (MA): Addison-Wesley

Peirce, C.S. 1903: A Syllabus of Certain Topics of Logic, Collected Papers, 1.180-202, 2.219-225 and other paragraphs

Perrow, C. 1984: Normal Accidents. Living with High-Risk Technologies, New York: Basic Books

Polanyi, M. 1966: The Tacit Dimension, Garden City (NY): Doubleday

Prasse, M. & Rittgen, P. 1998: Bemerkungen zu Peter Wegners Ausführungen über Interaktion und Berechenbarkeit, Informatik-Spektrum 21, 141-146 (Remarks on Peter Wegner's statements about interaction and computability)

Putnam, H. 1960: Minds and Machines. In: Hook, S. (ed.): Dimensions of Mind, New York: Collier Books

Putnam, H. 1991: Representation and Reality, Cambridge (MA): MIT Press

Reckwitz, A. 2002: Toward a Theory of Social Practices: A Development in Culturalist Theorizing, European Journal of Social Theory 5 (2), 243-263

Ryle, G. 1949: The Concept of Mind, London: Hutchinson

Schmidhuber, J. 2015: Deep Learning in Neural Networks. An Overview, Neural Networks 61, 85–117

Searle, J.R. 2010: Making the Social World. The Structure of Human Civilization, Oxford: Oxford University Press

Shannon, C. 1948: A Mathematical Theory of Communication, The Bell Systems Technical Journal 27, 379-423 and 623-656

Sharif, M. et al. 2016: Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition, ACM Conference on Computer and Communication Security, Vienna

Shoham, Y. 1993: Agent-Oriented Programming, Artificial Intelligence 60, 51-92

Solow, R. 1987: We We'd Better Watch Out, Th New York Times Book Review, July 12

Spur, G. 1984: Über intelligente Maschinen und die Zukunft der Fabrik, Forschung – Mitteilungen der DFG, I-VIII (About intelligent machines and the future of the factory)

Varela, F.J.; Thompson, E. & Rosch, E., 1991: The Embodied Mind. Cognitive Science and Human Experience. Cambridge (MA): MIT Press

WEF 2012: The Future of Manufacturing. Opportunities to drive economic growth, a World Economic Forum Report in collaboration with Deloitte Touche Tohmatsu Limited

Wegner, P. 1997: Why Interaction is More Powerful than Algorithms, CACM 40 (5), 80-91

Wiener, N. 1950: The Human Use of Human Beings. Cybernetics and Society, Boston (MA): Houghton Mifflin Harcourt

Winograd, T. 1996: Bringing Design to Software, Reading (MA): Addison-Wesley

Wooldridge, M. 2002: An Introduction to Multi-Agent Systems, New York: Wiley