ARTIFICIAL INTELLIGENCE AND THE FUTURE OF MANAGEMENT[1]

STEWART CLEGG

MARCO BERTI

ACE SIMPSON

MIGUEL PINA E CUNHA

[1] This work was funded by National Funds through FCT – Fundação para a Ciência e Tecnologia under the project Ref. UID/ECO/00124/2013 and by POR Lisboa under the project LISBOA-01-0145-FEDER-007722

INTRODUCTION: AI AND ORGANIZATION

Humanism is on the retreat and the present cycle of managerial innovation sees a favouring of machine over human imagery (Abrahamsson and Eisenman, 2008) that has the potential of paradoxically being both as liberating for humanity as it is dehumanizing. What are the implications of these trending fashions and what possible directions might emerge from them? With the rapid advancements in Artificial Intelligence (AI), organisations are increasingly dealing with artificial actors whose intelligence is fundamentally different from human intelligence (Ayoub and Payne 2016), namely from human practical intelligence.

The concept of intelligence is difficult to capture (Wang, 2007): it has been variously defined as "the complex expression of a complex set of principles" (Yudkowsky, 2007: 389); the capability to meet challenges by transforming information into knowledge (Glynn, 1996); the "ability to achieve outcomes that fulfil desires" (March, 1999: 1); and "the ability to adapt with insufficient knowledge and resources" (Wang, 2007: 31). In the context of AI it is important to distinguish between "narrow AI" and "Artificial General Intelligence" (Pennachin & Goertzel, 2007: 1) or AGI. The former refers to a specialized capacity to perform clearly defined tasks, solving narrowly defined problems, as 'GOFAI - good old fashioned AI' (Haugeland, 1985). Based on a purely mathematical logic, this narrow AI is already well developed (Broussard, 2018), and characterizes "AI partners" (Etzioni & Etzioni, 2017: 411), instruments that provide smart assistance to human decision-makers. The latter is instead still a hypothetical capacity, which would involve "an ability to acquire and apply knowledge, and to reason and think, in a variety of domains" (Pennachin & Goertzel, 2007: 6). It implies the development of "AI minds" (Etzioni & Etzioni, 2017: 411) capable of passing the Turing test, i.e. of actions that are indistinguisheable from those of a

human agent. While this is already part of our imaginary, as the "Hollywood kind of AI" (Broussard, 2018: 32), it is still well beyond the reach of current technology.

Machine learning is seen as a necessary (but insufficient) element for attaining AGI. It is based on two operations (algorithms): a first which learns bottom-up from some training data (for instance records of decisions made by humans) and a second which uses the rules 'learnt' by the first system to classify new inputs. This process is opaque, in the sense that is not clear how a particular classification decision has been reached from inputs (Burrell, 2016). In this AI machines learn and solve problems on their own, by mimicking human cognitive functions. Every encounter with new data enhances the breadth and depth of the machine's internal schema categories, providing reference points for processing new incoming data using neural reference simulations, probabilistic inferences and pattern analysis.

By mid-2018 over a dozen countries had proposed new AI strategies (Dutton, 2018), while industry is developing AI principles and developing best practices. They are also involved in developing regulation for AI, whether through direct participation or lobbying efforts. As Cath (2018: 1) suggests, such industry efforts are laudable but imply important ethical questions: notably, who sets the agenda for AI governance; what cultural logic frames it, and who benefits?

Artificial Intelligence must coexist with human intelligence, whose practical wisdom it may supplement, challenge or even contest. We explore the direction of these developments, the legitimacy of AI's involvement in managing organisations and the role of managers and leaders in this changing context. More specifically we ask: what does the growth of AI systems in business organisations imply for the way that managers will need to exercise human practical wisdom? By practical wisdom we refer to "the application to professional

pursuits of a deep understanding and fundamental capacity for living well" (Kessler & Bailey, 2007: lxvii) or the capacity to "bridge contradictions" (van Dierendonck & Patterson, 2015: 123). There are many contradictions at the interface of human and AI, as we elaborate next, priot to reflecting on how phronesis might be cultivated to address the tensions raised at the interface of forms of human and AI.

UNDERSTANDING MACHINES

Human intelligence concerns more than just abstract information processing. It is situated, value-laden and deeply rooted in the emotional life of the human beings that express it, implying knowing and knowing how to use knowledge (Tsoukas, 2017). In the field of organization studies, researchers are struggling to understand what it means to use AI knowledge wisely.

When discussing AI and the future of management it is tempting to speculate about scenarios in which AGI has been attained. Drawing on fictional examples is not the focus of this chapter: there are many famous insatnces that one could discuss, such as Hal in 2*001: A Space Odyssey* or the ethical challenges posed by the sentient robots described by Asimov, to speculate on the possibilities of developing intelligent machines that incorporate practical wisdom (see for instance Bryson, 2010; Gerdes & Øhrstrøm, 2013). While science fiction offers a useful device for understanding contemporary fears and concerns, as they are projected in a speculative universe, it is of less scientific use in predicting the future. Imagining AGI as a current reality may help with identifying sources of moral panic, while providing material for interesting thought experiments; however, it tends to exagerate machine capabilities and overlooks the challenges posed by the diffusion of current technologies. Indeed, real AI capabilities are quite limited, thwarting any dream of "technochauvinism… the belief that tech is always the solution" (Broussard, 2018: 7-8).

As our intent is to reflect on the implications of current AI technology for management, we intentionally avoid antroporphomizing machines or attributing them with thinking and feeling capabilities. On the other hand we are also careful not to treat artificial constructs as passive, docile entities at the service of their human masters. AI is an actant, i.e. "an entity endowed with the ability to act" (Callon, 1995: 53), that is, to make a difference (Latour, 2005) regardless of intentions or motivations, which can only be ever socially ascribed anyway (Blum and McHugh, 1971). Technology can therefore be understood "as a family of methods for associating and channelling other entities, both human and nonhuman" (Law, 1987: 233) and understood on the basis of its actual performativity (Latour, 2005). From this perspective it is important to consider human actors and non-human AI symmetricaly (Callon, 1986) if only, irrespective of their actual 'intelligence', because their actions have relevant consequences.

AI is endowed with agency; consequentially it has management implications. We can consider these through the relevance of terminological differences in relation to managing the ethical consequences of the diffusion of AI in contemporary organizations. Having already mentioned the important distinction between general and narrow AI, another terminological conseration is the entity pictured as the embodiment of AI. Two alternatives are frequently discussed: robots (Bryson, 2010; Gerdes, 2016; Sharkey, 2017) and algorithms (Martin, 2018; Ziewitz, 2016). The problem with the former is the risk of anthropomorphising the issue, taking for granted that the final objective of AI development is building an artificial human, forgetting that real robots are almost always highly specialized and limited machines. With the latter, discussions are often dominated by a misleading narrative of an "algorithmic drama" (Ziewitz, 2016: 5), representing them as powerful and inscrutable artificial entities, overlooking the fact that they are indeed ambiguous, emergent, and hybrid.

Beyond the emotional and rhetorical effects deriving from the embodiment of AI, application stimulates rich debate concerning the possibility of creating artificial moral agents (van Wynsberghe & Robbins, 2018). Here two separate issues are considered. The first concerns technical feasibility, i.e. the actual possibility of creating a morally intelligent machine. Different approaches considered include those based on top-down normative principles, bottom-up machine learning, or a hybrid (Gerdes & Øhrstrøm, 2013). A lack of progress in AGI hinders success in this field as robots lack capacities for feeling and desiring, requisites of moral action (Gerdes, 2016). Revisting traditional methods used for controling human behaviour, such as general laws and rules, while remaining conscious of their limitations (Etzioni & Etzioni, 2017), is required. Some 'technochauvinists' claim that machines would outperform humans as moral agents since they are not biased by interests and emotions (Dietrich, 2001), forgetting that "humans do sometimes make flawed decisions, but [unlike machines] they can reflect and learn from them" (Sharkey, 2017).

The second aspect of the discussion concerns the desirability of providing machines with ethical intelligence. An examination of the justifications in favour of developing moral reasoning capabilities in AI reveals multiple flaws (van Wynsberghe & Robbins, 2018). AI can be simply be used to augment human judgment capability, instead of replacing it; harm prevention can be guaranteed by 'dumb' safety features; the complexity of a sufficiently intelligent machine precludes transparency (and vice versa) etcetera. A central aspect is the often misunderstood notion of *autonomy,* which is a loaded term, since we typically associate it with freedom of choice and individual rights. *Machine autonomy*, however, intended as the capacity to independently adapt to unforseen circumstances while performing a task, is not the same as *moral autonomy*, which implies the capacity to choose (and ethically evaluate) objectives (Etzioni & Etzioni, 2017). A reflection on autonomy also has important implications for the use of algorithms: while there is consensus on their increased influence

on decision-making, the fundamental responsibilities of the companies developing them tend to be overshadowed by a "fase tension" between their representation as "objective… versus deterministic… agents" (Martin, 2018: 2).

If AI is really much more limited and 'slavish' (Bryson, 2010) than most realise, it is necessary to consider the ethical and political implications of its increased use in relation to its human masters, considering how it concretely changes organization. Rather than fearing hypothetical dangers, such as the effects of technological singularity (Eden, Moor, Søraker, & Steinhart, 2012), we must reflect on how the mis-use of these not-so-intelligent agents can produce undesirable effects. An example is the social isolation effects of using robots to mind vulnerable individuals, such as children or the elderly (Sharkey, 2008): avoiding these shortfalls does not require the development of superintelligent moral machine but wise judgment, together with legislation to keep in check human greed and selfishness (Etzioni & Etzioni, 2017). We need to examine technology in an organizational context, while recognising that we might have already developed a form of artificial general intelligence: *organization*.

ORGANIZATIONAL INTELLIGENCE AND TECHNOLOGICAL BLACK-BOXING

AI already exists, it can be argued, embodied in organizations rather than in robots or computers, as an artificial entity (i.e., a system designed by humans) endowed with the capacity "to adapt to its environment while operating with insufficient knowledge and resources" (Wang, 2007: 33). As a working definition of intelligence that has been devised in the AI field this clearly resembles traditional definitions of organization as open systems (Boje, 2008); thus, it connects with an idea of organizational intelligence (OI) already half a century old (Wilensky, 1967). Striking similarity is found between the description of AI

presented above and the most influential definition of OI, as "an organization's capability to process, interpret, encode, manipulate, and access information in a purposeful, goal-directed manner, so it can increase its adaptive potential in the environment in which it operates" (Glynn, 1996: 1008). In other words, OI represents "the capacity to adopt procedure that consistently do well (in the organization's own term) in the face of constraints" (March, 1999: 1). Such an adaptive capacity is not just the product of the accumulation of individual members' intelligence but includes also the transformation and codification of individual intelligence and their embedding in structured patterns of thought and action (routines, norms, procedures etc.) (Glynn, 1996).

The introduction of semi-autonomous machines does not qualitatively transform this assemblage of human and technology any more than the introduction of a new routine or non-adaptive technology (e.g. a conveyor belt) transforms them: the operational, political and ethical challenges of organizing remain the same. A routine, which includes both a normative element as well as adaptive improvisation on the part of its performers (Feldman & Pentland, 2003; Pentland & Rueter, 1994), is not dissimilar to a set of machine learning algorithms. Indeed, what algorithms do most easily is to capture and replicate routines; that is why algorithms increasingly replace human labour in areas such as booking flights, hotels, or searching for basic information. Algorithms accomplish laborious routine tasks better than slower and less consistent and predictable humans.

A first concrete ethical issue emerging from the use of AI in organizations that has nothing to deal with maverick robots. Rather it concerns the question of what happens to the people whose jobs are displaced by AI and who sets the agenda for their futures? Frey and Osborne (2017) argue that half the jobs in the US economy were likely to be eliminated by algorithms for big data based upon pattern recognition that can readily substitute for labour

in a wide range of non-routine cognitive tasks (Brynjolfsson & McAfee, 2012; Manyika et al., 2013). Combined with the fact that advanced robots are gaining sensemaking capabilities and manual dexterity (Manyika et al., 2013; VO, 2013), the nature of work across industries and occupations are likely to change dramatically. On this basis they identified nearly 50% of existing jobs as under threat of routinization and disappearance within a decade. As many of the industry leaders in the field of AI are organizations that are domiciled in the United States, it is important to recognize the extent to which instititional logics embedded in US practices might implicitly shape the field, especially in terms of issues of skill formation, training and re-training. In the US these tend to be for more voluntarsistic than in the most progressive social democracies.

The modelling conducted by Frey and Osborne (2017) predicted that most workers in transportation and logistics occupations, the bulk of office and administrative support workers, labour in production occupations, and a substantial share of employment in service occupations, site of most recent US job growth (David & Dorn, 2013), are highly susceptible to computerisation and the growth in the market for service robots (Manyika et al., 2013) and the concomitant gradual diminution any comparative advantage human labour might have in tasks involving mobility and dexterity (VO, 2013). Displaced workers will have to participate in education and training that boosts their skill formation or risk falling by the wayside, as labour becomes surplus to contemporary requirements. Whether such education and training is left to market foces or the role of the state is a vital issue for those whose livelihoods are under threat. Yet, the assumption that all individuals have unlimited potential and are 'reprogrammable' through 'retraining' is problematic. Intrinsic differences in intellectual capabilities and the insurmountable disadvantages determined by social inequities mean that not every employee will turn into a 'creative economy knowledge worker'. Furthermore, merely casual work is often marketed as entrepreneurial

opportunity. Indeed, the freedoms provided by technological advances can paradoxically be akin to slavery (Dubal, 2017).

Even if the game has not changed (the theme of job loss to AI clearly echoes the Luddite concerns about industry mechanization in the 19th century), the massive introduction of AI in organizations nontheless has implications beyond direct competition between human and artificial labour. The impacts can be better understood by reflecting on organizational phenomenon as not just a 'thing', a mechanistic assembly of technology and people, but as a process, the contingent outcome of organising performances (Clegg, Kornberger, & Rhodes, 2005; Czarniawska, 2008), built upon sensemaking acts that enable interlocking behaviours (Weick, 1979). Multiple agencies, both human and non-human, are implicated in this constant laboring, in which "in the thick of being an organizer, it is utterly impossible to distinguish organization and disorganization" (Latour, 2012: 166-167). Organizations are therefore panoplies of flexible entities whose mutually adjusting aims at producing action (Çalışkan & Callon, 2010). Both symbolic and material elements are implicated in this process, a perspective from which it is possible to consider the implications of the extensive use of algorithms and robots in organizing.

As previously mentioned, both algorithms and robots tend to be rhetorically misrepresented as powerful, autonomous agencies (Bryson, 2010; Martin, 2018; Ziewitz, 2016). As a misrepresentation, this is not problematic because of the risk of a 'HAL-syndrome' but because it becomes a device for reducing corporate responsibily for the consequences of organizational actions. Algorithms make decisions based on an internal logic that is not transparent (Luca, Kleinberg & Mullainathan, 2016); yet designing (and using) an algorithm implies a process of delegating decisional responsibilities, and delegation does not remove responsibily (Latour, 1992).

Reducing the issue of responsibility to one of technical choice represents a fundamental flaw in 'managing with AI'. Technology is constituted within an ensemble of social, political, economic relations, and exists within a certain mode and relations of production (Marcuse, 2004). In the case of algorithms this happens both by accepting their black-boxing (as imprescrutable, self generating entities) and by considering their control a purely technological issue. From this perspective they represent the purest form of technically rational management. Such forms of management are not confined to the workplace but are characteristic of socialization practices within the home and at school premised on increased AI surveillance and interaction based on the pre-coded recognition of behavioural signs (Chan, 2018; Finkel, 2018).

A further illustration of these issue occurs in the debate on self-driving vehicles. For instance, the German government has issued rules for the programming of autonomous vehicles with the intent of making sure that machine intelligence will comply with duty of care principles (The Federal Government, 2017). Acknowleding the limitations of the judgment of machine intelligence, these principles prescribe that, in case of an unavoidable collision, the AI should opt for harm minimization making no discrimination on the basis of person's relative worth (e.g. in terms of age, gender, health, relationships etc.). While these normative principles seem reasonable, they demonstrate the perils of considering the management of unpredictable, morally charged situations as a technical problem rather than one that is sociological. First, this principle overlooks that people have inconsistent requests, in that most agree with driverless cars making utilitarian decisions but only if these decisions do not pose a risk to the car's occupants (Bonnefon, Shariff, & Rahwan, 2016): in practice, no one would buy a car that decides to kill the driver to save pedestrians (which makes the issue moot). Second, it is a mistake to draw on extreme scenarios (such as the 'trolley problem'), because they are very rare and based on extreme simplification. The problems encountered

by a self-driving cars are the same as those encountered by a human driver: in most cases choices are standardizable and are regulated by legal dictates (e.g. traffic rules) that are formulated through collective decision making. In following such rules a machine can actually be more consistent (and law abiding) than a human driver, minimizing accidents. In unpredictable and sudden events requiring moral deliberation (saving the driver or another person) it could be argued that the choice is best made randomly, similar to the instinctive, pre-rational deliberation that would be made by a human driver in those circumstances (Etzioni & Etzioni, 2017).

These examples illustrate that some of the issues associated with AI are sociological (implying political, discursive and ethical components), rather than merely technical. Use of AI in organizations is only going to increase into the future. Accordingly, the future of management requires addressing the tenions emergent from human and AI interactions. As many of these tensions are paradoxical in nature, there is a need for developing competencies in practical paradoxical wisdom (Rodrigues et al., 2017).

WORKING WITH AI

With the advance of organizational AI use, much of management will entail coordinating those who frequently work with AI, knowledge workers. An increase in knowledge-intensive work means organizations have to employ – and manage – different kinds of employees. Mental rather than manual labour, is the order of the day. Employees need to be capable of working with sophisticated databases, software and knowledge management systems. These have to be related to customer and client requirements often on a unique and tailored basis that deploys a common platform while customizing it for specific requirements. Thus, technical and relational skills will be at a premium, which implies a particular ethical approach to AI.

Advanced computer-assisted information processing technologies will increasingly result in the reduction of human nodes in an information network, leading to flatter organizations (Huber, 1990). Kallinikos (2007) argues that digital technologies allow tasks that were previously embedded in the 'fixed space' of traditional organizations (for example accounting, inventory management, production operations or financial management) to be dissolved and recomposed as 'informatised' modules or services (Kallinikos, 2007: 96). Digital technologies are implicated in an historic shift dissolving bureaucratic organization, such as Zappos' holacracy circles (Greenfield, 2015). The major advantage of digital technologies for business and organizations are their virtual possibilities for disaggregating existing designs. Increasingly, organizations will specialize in activities that are critical to their competitive advantage and outsource elsewhere those that are not. Non-core functions, such as back office accounting, telemarketing, or programming, outsourced to parts of the world where the wage is one third to one tenth the cost in the home market, dramatically reduces operating costs and increases competitiveness (Davis, 2009). Accordingly, organizational efficiency, paradoxically, de-materializes, deconstructs and fragments that which it makes efficient.

Digital affordances will mean that distinct organizations can work together, provided there is trust, empathy and commitment on all sides. Control over data "will be distributed across complex socio-material networks that include interested organizations, automated algorithms, analysts and privately controlled meta-data systems" (Plesner & Gulbrandsen, 2015: 157). Boundaries, choices and control are all shifting in the direction of increasing fluidity and plurality. Paradoxically the openness and accessibility that is the strength of digital platforms is also their weakness, making them vulnerable to highjacking by those threatened by a free press, the contestation of ideas and liberal democratic values.

It is furthermore evident that workers who become bereft of their digital devices and their affordances will feel less than fully human – McLuhan's (1994 [1964]) hypothesis that the media become extensions of our nervous systems holds even more so than when first formulated. The nature of being an individual is changing as anyone can be connected anywhere, anytime through social media. On the one hand, social media can build rapid momentum in mobilizing blocs of voters or consumers; on the other hand, it can be used as a means of distraction and appeasement. Digital devices increasingly provide membership of virtual communities; networking proclivities, interests and desires; creating parallel digital bubbles which workers increasingly inhabit (Fosfuri, Giarratana, & Roca, 2011).

The social organization emerging from this technologically mediated network of relations is a cyborg-organization, partly human and partly machine. If IT enhanced social network are emergent entities, other organizations have purposefully turned into cyberfactories (Czarniawska, 2011). In a sense, AI is already around us, not as a silicon-based electronic brain but as a hybrid product of human-machine integration. It appears AI is not just created to 'liberate' workers from tedious jobs or to replace them with more docile and resilient robots but also the better to control and guide cyber-enhanced and, perhaps ultimately, cyborg (Haraway, 1991) employees (e.g. the Amazon.com patent on haptic bracelets to guide their warehouse worker), deprived of any need for practical wisdom and critical thinking. It is cheaper to 'augment' cheap flesh and blood employees rather than develop very expensive autonomous robots. Accordingly forced cyberization will increasingly affect the workforce. The owners of machines, in making work lighter through cooperation between machines and humans, paradoxically appropriate increased workloads (while destroying other jobs): the paradox of exploitation, explored at length in Marx's *Capital* (1976).

FROM QUESTIONS OF INTELLIGENCE TO CONCERNS ABOUT ETHICS

Most discussion on the issue of 'artificial moral agency' and on the ethical consequences of using AI imply a normative/cognitive ethical understanding, based on the assumption that "ethical behavior is primarily a cognitive process" (Treviño, Weaver, & Reynolds, 2006: 979) dependent on moral reasoning capabilities (Kohlberg, 1981). Consequently, considerings ethics as "the adjudication of what to do in complex ethical choices" (de los Reyes, Kim, & Weaver, 2017: 315), implies understanding moral decisions as the logical and necessary outcome of a set of abstract deontological principles defining how 'one ought to act'. This view implies considering ethical decision-making as a staged process, wherein ethical behavior is the final step in a process that includes moral awareness, moral judgment and ethical intentions (Rest, 1986). The model can be enriched to consider contextual barriers to the transformation of cognition in action (Hannah, Avolio, & May, 2011) but ultimatelty all variations on this theme consider ethical decisions as a deliberate rational process. In such cases, an ethical approach is aligned with an 'algorithmic' view that evaluates situations and behaviors by processing information on the basis of a defined set of normative instructions. Typically, when discussing issues surrounding the introduction of AI the focus (as in the German driverless car case) converges on a normative design problem (how to decide the 'correct' set of rules that will guide machine decisions). As discussed before, this leads to an intractable problem, due to the absence of an agreed upon "moral epistemology" (van Wynsberghe & Robbins, 2018: 12).

A complementary way of considering the moral decision problem is offered by behavioral ethics (Treviño et al., 2006). While rationalist models view emotions as a possible input to reasoning but never as a direct cause of moral judgments, behavioral ethicists are concerned with understanding the actual basis of observed moral behavior (why individuals act in a certain way). Informed and inspired by neurocognitive studies and behavioral economics (Kahneman, 2011; Lieberman, Gaunt, Gilbert, & Trope, 2002), this perspective identifies

two interrelated operating systems of 'moral reasoning': one that is reflexive/intuitive, based on pre-conscious schema, while the other is reflective and deliberative. These dual mechanisms work better or worse depending upon the context: moral intuitions are reliable in micro decision contexts, while reasoning is more more effective in impersonal macro contexts (Orlitzky, 2017). Automatic processes represent devices for a quick and efficient means of guiding behavior in routine situations (Reynolds, 2006), while emotions such as guilt (Tangney, Stuewig, & Mashek, 2007), gratitude (McCullough, Kilpatrick, Emmons, & Larson, 2001) or disgust (Schnall, Haidt, Clore, & Jordan, 2008; Tangney et al., 2007) play a central role in triggering moral responses.

The higher-order reflective system can be used for moderating 'gut-feeling' based decisions (Pizarro & Bloom, 2003) as well as for adjusting the prototypes on which information is structured (Reynolds, 2006). Conversely, 'motivated reasoning' can operate as a rationalization device (Haidt, 2001). Ethical intentions and ethical behavior, which appear as separate stages in a cognitive model, are in reality intertwined (Reynolds, 2006). Mental maps play a central role in constructing what is identified as 'facts' (Bateson, 1979; Weick, 1995): hence the information processed by the higher-level conscious reflective system is unavoidably be filtered by the intuitive system.

One might think that delegating decisions to AI is preferable, since with an algorithm 'moral reasoning' cannot be affected by invisible cognitive biases (any bias should be visible in the structure of the algorithm); nor would its action be affected by lack of moral efficacy (belief in one's capacities to act in response to the issue) or moral courage (the will to face threats deriving from acting) (Hannah et al., 2011). As such, AI could warn humans of the moral implications of actions, helping avert phenomena such as moral muteness, the reluctance to describe actions in moral terms even when they are clearly of moral nature (Bird & Waters,

1989), 1989) or moral blindness, the temporary inability to see the ethical dimension of a decision (Palazzo, Krings, & Hoffrage, 2012).

Yet the assumption that there is an absence of bias is illusory: it is not correct that AI systems are not affected by cognitive biases. Machine learning systems learn biases (invisibly) from their training data (Angwin et al., 2016; Mooney, 1996). For example, it has been shown that algorithms can actually be racist at least in a performative sense (Angwin, Larson, Mattu, & Kirchner, 2016). Unconscious biases (often linked to gender, class and race) influence programmers (Broussard, 2018). Since they are not incorporated in explicit decision rules such biases cannot be easily tracked because they can be introduced through skewed choices of training and source data (Martin, 2018). These considerations open a number of very important questions and themes useful for focusing future research. For instance: how can human/AI decisional interfaces be designed, given the role of intuitive moral behavior? How can the human-machine interface be designed so it is possible, at a social/organization level, to leverage their specific weaknesses and strengths?

As described earlier, a lack of sentience and self-awareness inhibit 'intelligent' machines from assessing the situated implications of decisions, and making reflective decisions with a sense of personal responsibility. Their judgment calls are therefore constrained by a rule based ethics reasoning which leads to paradoxes and problems deriving from the attempt to apply abstract universal norms to a complex and nuanced reality. Therefore, as the quantity and relevance of decisions delegated to AI devices increases, to a corresponding degree 'phronetic supervision' that gives consideration to the impact of decisions on different stakeholders and the desirability such outcomes (Flyvbjerg, Landman, & Schram, 2012) needs to increase as well. The practical implication of this line of reasoning is that business education and executive training must increasingly devote energies towards developing

phronetic capabilities of its students, rather than merely increasing their technical literacy. For many this is not obvious and goes against the grain of the fashionable 'STEM' focus that characterizes contemporary academia.

ETHICS AND POLITICS OF AI

Considered in the broader workplace context, these challenges are suggestive of the need to cultivate phronetic leadership. As Shotter and Tsoukas (2014: 224) put it, "Phronetic leaders are people who, in their search for a way out of their difficulties, have developed a refined capacity to intuitively grasp salient features of ambiguous situations and to constitute a landscape of possible paths of response". We propose some basic guiding principles that can inform the cultivation of phronetic capabilities within the workplace context. Protection of human beings from technology of a different type, medicine, is the reason why in the third century BC Hippocrates penned an oath to which doctors are still held accountable today. We suggest that these standards can inform the development of phronetic leadership for management of AI within organizations. First, AI technology must not cause harm to human workers. Second, all technology must be used for the good of the workers, with no human deprived of their fundamental rights by a computer, and when there is a question about how technology affects fellow humans, that is a decision that should only be made with human consent. Third, AI technology must always be used in a manner that respects human privacy, so that the relationships between humans remains as a relationship of trust. Conceptions of human privacy are extremely variable across societies; those dominat in the United States, where most of the technology originates at present, may not be appropriate elsewhere in jurisdictions with stronger privacy laws, as in much of Europe. As Cath (2018: 4) argues, "An American, corporate needs-driven agenda is not naturally going to be a good fit for the rest of the world".

In pursuit of the objective of doing no harm, it should remembered that basic issues of exploitation are rarely far from most employment relations. AI, potentially, could precipitate the crisis of capitalism that Marxists have long anticipated and markets have long feared. As living humans are replaced by robots and intelligent machines the risk is that as the global reserve army of the unemployed and under-employed swells, proletarian immiseration will lead to a growth in collective consciousness concerning the consequences of existing relations of production. Specific managers of specific organizations, heeding the necessity of feeding the voracious appetites of shareholder value, may well rapidly replace living labour power where they are able to with the machine power of AI, heedless of the concern that labour that produces also consumes and reproduces other consumers and that in order to maintain collective consumption their needs to be aggregate effective demand.

If the option of shifting to machine-based labour through AI and roboticization is one concern for the future of the advanced capitalist societies through marginalizing human labour, paradoxically there is an equivalent ethical concern in not doing so. Resisting using AI capabilities and preferring to exploit human labour as effectively as possible through maintaining low-wage, low-tech, low-productivity organizations should not, ethically, be a feasible alternative for employers. Employers should not be able to profit from the infliction of low productivity jobs on workers on low wages and it should be the role of the state to regulate the market to ensure that they do not do so. Short of the low wages option, with effective national training schemes in place that enable employees to become adept at working with AI, employment should shift into higher productivity areas of the economy. Workers that became unemployed in some sectors should be trained to become more productive workers, rather than left as the flotsam and jetsam of ‘market forces’. Real wages and effective demand would rise, as would profits which, to the extent that they were re-invested in improving the productive capacity of corporations, could be favourably taxed as

incentives that favour long-term investments in research and development, rather than capital gains and dividends. Those that continued to feed the voracious appetites alluded to in preference to innovation and accompanying skill formation could enjoy a more burdensome form of regulation and taxation. AI thus raises ethico-political issues as well as strictly ethical considerations.

In pursuit of the objective of using technology for the good of workers, with no human deprived of their fundamental rights, the objective of policy should be to make labour gradually more expensive, even labour whose primary social function is its supply of cheap services. More expensive labour is not only a personal good for those that receive it but is also an institutional good in building social cohesion. In the face of rising costs, the objective should be for work that exists only because of the low cost of labour either to be outsourced elsewhere as employers increasingly adopt robotics and AI outsourcing increases distance and potentially decreases practical wisdom and critical reasoning, as it may lead organizations to represent laborers as sources of efficiency rather than of improvement and to design jobs poorly (Ton, 2012).

Bill Gates has suggested that productivity-enhancing devices should be taxed in the developed economies. The main argument against taxing robots is that it might impede innovation but this is the case only if the option of a low-cost workforce is available: if the opportunities for that are limited through wages legislation and enforcement the objections would be dissolved. Enterprises that in the past relied on super-exploitation of low-cost labour would either die as they were priced out of the market or would adjust structurally through increasing investment in productivity enhancements that would improve working conditions, wages and productivity. To the extent that industries decamp elsewhere, all well and good: they will, in all probability, lift average wages where they land and, provided there

is an active labour market policy funded in part by an exit tax on de-campers, then the overall level of skill formation and domestic income can be raised.

Finally, in pursuit of the objective of using AI technology only in a manner that respects human privacy, employee data must be collected and analysed under strict conditions. AI applications for financial control, automation of professional as well as manual jobs, surveillance of employees, performance management, marketing and customer relation systems must not invade employee privacy. Employees must be made fully aware of what data is being collected, how it is to be used and they must have given their prior consent. Under no circumstances should employers sell such data to third parties for commercial gain. A further consideration here is the delegation and the locus of responsibility – what does it mean to delegate a decision to a machine, and does this confuse lines of accountability? Wise management should accordingly ensure that AI technology does not affect humans without their consent.

FINAL COMMENT

The paradoxical sense of a beginning that also spells an end is palpable: organizations seeking to simplify and order work are making it more complicated by replacing people with complex machines; innovation is creating advantages that are also destroying social systems; open digital information systems that provide for democratic pluralism are being weaponised by those who feel threatened by plurality of voice; freedoms provided by technology are causing greater enslavement; all the while management as a human practice is increasingly becoming a machine practice. The implications for the future of management pose wider questions for the future of economies and societies. There is the large and interesting question of the role of organizational leadership in determining the technological future of management. Technology does not evolve. It is developed by people and, most notably, by

organisations. Wise leadership should have something to say about what kinds of technology should be developed or not and how it is used ethically within organiations. Wise leadership would also not merely leave these matters to the market or organizations to determine what practices are institutionalized. It is important to ensure that there is representation of civil society concerns when regulating AI, as well as a realization that there is a need for more non-US led initiatives such as the Europe-based AI4People (http://www.eismd.eu/ai4people/) and the Council on Europe's Expert Committee on AI and Human Rights (https://www.coe.int/en/web/freedom-expression/msi-aut). Not only government, and industry but also civil society leadership are required.

REFERENCES

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. 2016. Machine bias: There's software used across the country to predict future criminals. *And it's biased against blacks.* ***ProPublica***.

Bateson, G. 1979. ***Mind and nature: A necessary unity***. New York: Dutton.

Bird, F. B., & Waters, J. A. 1989. The moral muteness of managers. ***California Management Review***, 32(1): 73-88.

Blum, A. F., & McHugh, P. (1971). The social ascription of motives. *American Sociological Review*, 98-109.

Boje, D. M. 2008. Open systems. In S. R. Clegg, & J. R. Bailey (Eds.), ***International encyclopedia of organization studies***, Vol. 1: 998-1000. Thousand Oaks, CA: Sage.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. 2016. The social dilemma of autonomous vehicles. ***Science***, 352(6293): 1573-1576.

Broussard, M. 2018. ***Artificial unintelligence: How computers misunderstand the world***. Cambridge, MA: MIT Press.

Brynjolfsson, E., & McAfee, A. 2012. ***Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy***.Boston, MA: The MIT Center for Digital Business.

Bryson, J. J. 2010. Robots should be slaves. In Y. Wilks (Ed.), ***Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues***: 63-74. Amsterdam: John Benjamins.

Burrell, J. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. ***Big Data & Society***, 3(1): 2053951715622512.

Çalışkan, K., & Callon, M. 2010. Economization, part 2: a research programme for the study of markets. ***Economy and Society***, 39(1): 1-32.

Callon, M. 1986. Some Elements of a Sociology of Translation-Domestication of the Scallops and the Fishermen of St-Brieuc Bay. In J. Law (Ed.), ***Power, Action and Belief: A New Sociology of Knowledge?***: 196-223. London: Routledge.

Callon, M. 1995. Four models for the dynamics of science. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), ***Handbook of Science and Technology Studies***: 29-63. Thousand Oaks, CA: Sage.

Cath, C. (2018) Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A* 376: 20180080. http://dx.doi.org/10.1098/rsta.2018.0080

Chan, T. F. 2018. A school in China is monitoring students with facial recognition technology that scans the classroom every 30 seconds, ***Business Insider***.

Clegg, S. R., Kornberger, M., & Rhodes, C. 2005. Learning/Becoming/Organizing. ***Organization***, 12: 147-167.

Czarniawska, B. 2008. Organizations as obstacles to organizing. ***Nobel Symposium Foundations of Organization August 28-30, 2008 Stockholm***.

Czarniawska, B. 2011. ***Cyberfactories: How news agencies produce news***. Northampton, MA: Edward Elgar.

David, H., & Dorn, D. 2013. The growth of low-skill service jobs and the polarization of the US labor market. ***American Economic Review***, 103(5): 1553-1597.

Davis, G. F. 2009. The rise and fall of finance and the end of the society of organizations. ***Academy of Management Perspectives***, 23(3): 27-44.

de los Reyes, G., Kim, T. W., & Weaver, G. R. 2017. Teaching ethics in business schools: A conversation on disciplinary differences, academic provincialism, and the case

for integrated pedagogy. ***Academy of Management Learning & Education***, 16(2): 314-336.

Dietrich, E. 2001. Homo sapiens 2.0: why we should build the better robots of our nature. ***Journal of Experimental & Theoretical Artificial Intelligence***, 13(4): 323-328.

Dubal, V.B., 2017. ***Wage Slave or Entrepreneur: Contesting the Dualism of Legal Worker Identities.*** *California Law Review*, 105(1), 65-123.

Dutton T. 2018 Politics of AI, an overview of national AI strategies. See https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd.

Eden, A. H., Moor, J. H., Søraker, J. H., & Steinhart, E. (Eds.). 2012. ***Singularity Hypotheses: A Scientific and Philosophical Assessment***. Dordrecht: Springer.

Etzioni, A., & Etzioni, O. 2017. Incorporating ethics into Artificial Intelligence. ***The Journal of Ethics***, 21(4): 403-418.

Feldman, M. S., & Pentland, B. T. 2003. Reconceptualizing Organizational Routines as a Source of Flexibility and Change. ***Administrative Science Quarterly***, 48: 94-118.

Flyvbjerg, B., Landman, T., & Schram, S. 2012. ***Real Social Science: Applied Phronesis***. Cambridge, UK: Cambridge University Press.

Fosfuri, A., Giarratana, M. S., & Roca, E. 2011. Community-focused strategies. ***Strategic Organization***, 9(3): 222-239.

Frey, C. B., & Osborne, M. A. 2017. The future of employment: how susceptible are jobs to computerisation? ***Technological forecasting and social change***, 114: 254-280.

Gerdes, A. 2016. The Role of Phronesis in Robot Ethics. In J. Seibt, M. Norskov, & S. Schack Andersen (Eds.), ***What Social Robots Can and Should Do: Proceedings of Robophilosophy 2016/TRANSOR 2016***: 129-136. Amsterdam: IOS Press.

Gerdes, A., & Øhrstrøm, P. 2013. Preliminary reflections on a moral Turing test, ***Proceedings of ETHICOMP 2013, The Possibilities of Ethical ICT***: 167-174.

Glynn, M. A. 1996. Innovative genius: A framework for relating individual and organizational intelligences to innovation. ***Academy of Management Review***, 21(4): 1081-1111.

Greenfield, R. 2015. Zappos CEO Tony Hsieh: Adopt Holacracy or Leave. ***Fast Company***.

Haidt, J. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. ***Psychological Review***, 108(4): 814.

Hannah, S. T., Avolio, B. J., & May, D. R. 2011. Moral maturation and moral conation: A capacity approach to explaining moral thought and action. ***Academy of Management Review***, 36(4): 663-685.

Haraway, D. 1991. Simians, cyborgs, and women. New York: Routledge.

Haugeland, J. 1985. ***Artificial Intelligence: The Very Idea***. Cambridge, MA: MIT Press.

Huber, G. P. 1990. A Theory of the Effects of Advanced Information Technologies on Organizational Design, Intelligence, and Decision Making. ***Academy of Management Review***, 15(1): 47-71.

Kahneman, D. 2011. ***Thinking, fast and slow***. New York: Farrar, Straus and Giroux.

Kallinikos, J. 2007. ***The consequences of information: Institutional implications of technological change***. Cheltenham: Edward Elgar.

Kessler, E. H., & Bailey, J. R. 2007. Introduction: Understanding, applying, and developing organizational and managerial wisdom. In E. H. Kessler, & J. R. Bailey (Eds.), ***Handbook of organizational and managerial wisdom***: xv-lxxiv. Thousand Oaks, CA: Sage Publications.

Kohlberg, L. 1981. ***Essays on Moral Development, Vol. I: The Philosophy of Moral Development***. San Francisco, CA: Harper & Row.

Latour, B. 1992. Where are the missing masses? The sociology of a few mundane artifacts. In W. Bijker, & J. Law (Eds.), ***Shaping technology/building society: Studies in sociotechnical change***: 225–258. Cambridge, MA: MIT Press.

Latour, B. 2005. ***Reassembling the social: an introduction to actor-network-theory***. Oxford: Clarendon.

Latour, B. 2012. 'What's the Story?' Organizing as a Mode of Existence. In J.-H. Passoth, B. Peuker, & M. Scillmeier (Eds.), ***Agency without Actors? New Approaches to Collective Action*** 163–177. Abingdon: Routledge.

Law, J. 1987. On the Social Explanation of Technical Change: The Case of the Portuguese Maritime Expansion. ***Technology and Culture***, 28: 227-252.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. 2002. Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In M. P. Zanna (Ed.), ***Advances in experimental social psychology***, Vol. 34: 199-249. New York, NY: Academic Press.

Luca, M., Kleinberg, J., & Mullainathan, S. 2016. Algorithms need managers, too. ***Harvard Business Review***, 94(1): 20.

Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., & Marrs, A. 2013. ***Disruptive technologies: Advances that will transform life, business, and the global economy***. San Francisco, CA: McKinsey Global Institute.

March, J. G. 1999. ***The Pursuit of Organizational Intelligence: Decisions and Learning in Organizations***. Cambridge, MA: Blackwell.

Marcuse, H. 2004. ***Technology, War and Fascism: Collected Papers of Herbert Marcuse***: London: Routledge.

Martin, K. 2018. Ethical implications and accountability of algorithms. ***Journal of Business Ethics***: 1-16.

Marx, K. 1976. ***Capital: A critical analysis of Capitalistic Production***. London: Penguin.

McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. 2001. Is gratitude a moral affect? ***Psychological Bulletin***, 127(2): 249-266.

McLuhan, M. 1994 [1964]. ***Understanding media: The extensions of man***. Boston, MA: MIT Press.

Mooney, R. J. 1996. ***Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning.*** In Proceedings of the 1996 Conference on Empirical Methods in Natural Language Processing (EMNLP). 82–91.

Orlitzky, M. 2017. How cognitive neuroscience informs a subjectivist-evolutionary explanation of business ethics. ***Journal of Business Ethics***, 144(4): 717-732.

Palazzo, G., Krings, F., & Hoffrage, U. 2012. Ethical blindness. ***Journal of Business Ethics***, 109(3): 323-338.

Pennachin, C., & Goertzel, B. 2007. Contemporary Approaches to Artificial General Intelligence. In C. Pennachin, & B. Goertzel (Eds.), ***Artificial General Intelligence***: 1-29. Berlin: Springer.

Pentland, B. T., & Rueter, H. H. 1994. Organizational routines as grammars of action. ***Administrative Science Quarterly***: 484-510.

Pizarro, D. A., & Bloom, P. 2003. The intelligence of the moral intuitions: A comment on Haidt (2001). ***Psychological Review***, 110(1): 193-196.

Plesner, U., & Gulbrandsen, I. T. 2015. Strategy and new media: A research agenda. ***Strategic Organization***, 13(2): 153-162.

Rest, J. R. 1986. ***Moral development: Advances in research and theory***. New York: Praeger.

Reynolds, S. J. 2006. A neurocognitive model of the ethical decision-making process: implications for study and practice. ***Journal of Applied Psychology***, 91(4): 737-748.

Rodrigues, F.R., Cunha, M.P., Rego, A. and Clegg, S.R., 2017. The Seven Pillars of Paradoxical Organizational Wisdom: On the use Of Paradox as a Vehicle to Synthesize Knowledgeand Ignorance. In W. Küpers & O. Gunnlaugson, O. (Eds.), ***Wisdom Learning: Perspectives on Wising-up Business and Management Education*****.** Routledge, NY, NY.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. 2008. Disgust as embodied moral judgment. ***Personality and social psychology bulletin***, 34(8): 1096-1109.

Sharkey, A. 2017. Can robots be responsible moral agents? And why should we care? ***Connection Science***, 29(3): 210-216.

Sharkey, N. 2008. The ethical frontiers of robotics. ***Science***, 322(5909): 1800-1801.

Shotter, J., & Tsoukas, H. 2014. In search of phronesis: Leadership and the art of judgment. ***Academy of Management Learning & Education***, 13(2): 224-243.

Tangney, J. P., Stuewig, J., & Mashek, D. J. 2007. Moral emotions and moral behavior. ***Annual Review of Psychology***, 58: 345-372.

Ton, Z. 2012. ***The good jobs strategy: How the smartest companies invest in employees to lower costs and boost profits***. New York: Amazon Publishing.

Treviño, L. K., Weaver, G. R., & Reynolds, S. J. 2006. Behavioral ethics in organizations: A review. ***Journal of Management***, 32(6): 951-990.

Tsoukas, H. 2017. Don't Simplify, Complexify: From Disjunctive to Conjunctive Theorizing in Organization and Management Studies. ***Journal of Management Studies***, 54(2): 132-153.

van Dierendonck, D., & Patterson, K. 2015. Compassionate love as a cornerstone of servant leadership: An integration of previous theorizing and research. ***Journal of Business Ethics***, 128(1): 119-131.

van Wynsberghe, A., & Robbins, S. 2018. Critiquing the Reasons for Making Artificial Moral Agents. ***Science and Engineering Ethics***: 1-17.

VO, R. 2013. A Roadmap for US Robotics: From Internet to Robotics, 2013 edition. 23.

Wang, P. 2007. The Logic of Intelligence. In C. Pennachin, & B. Goertzel (Eds.), ***Artificial General Intelligence***: 31-62. Berlin: Springer.

Weick, K. E. 1979. ***The Social Psychology of Organizing*** (2 ed.). Reading, MA: Addison-Wesley.

Weick, K. E. 1995. ***Sensemaking in organizations***. Thousand Oaks, CA: Sage.

Wilensky, H. L. 1967. ***Organizational intelligence: Knowledge and policy in government and industry***. New York: Basic Books.

Ziewitz, M. 2016. Governing algorithms: Myth, mess, and methods. ***Science, Technology, & Human Values***, 41(1): 3-16.

AUTHORS' BIOS

Stewart Clegg is Distinguished Professor at the University of Technology Sydney and a Visiting Professor at Nova School of Business and Economics,. His research is driven by a fascination with power and theorizing. Stewart is a prolific writer and is the author or editor of a large number of books and journal articles and he often works on papers and books with his fellow authors with whom he worked on this contribution.

Marco Berti is Lecturer in Organization and Management in the UTS Business School, University of Technology Sydney, and a member of the Center for Business and Social Innovation. Marco's background is in management consulting and joined academia after an international career in organisational development. His main research interests are organisational paradox theory and the interaction between discourse and material space. Marco's research has been published in journals such as *Management Learning*, *Journal of Management Inquiry* and *The International Journal of Human Resources Management*. He authored the Elgar *Introduction to Organizational Discourse Analysis* (Edward Elgar**).**

Ace Simpson is Senior Lecturer in Organizational Behaviour in the UTS Business School, University of Technology Sydney, Australia and a member of the Centre for Business and Social Innovation. Ace has a strong interest in human well-being, flourishing and social justice, and his research is concerned with positive organisational practices such as humility, psychological safety, paradox transcendence and love. His main research focus is on the cultivation of organizational compassion, which he is currently studying as a critical missing factor in most programs aiming to address the persistent problem of workplace bullying. Ace's research has been published in journals such as the *Journal of Management, Journal of Business Ethics* and the *Journal of Management Inquiry*. Ace is also co-author of *Positive Organizational* Behavior (Routledge, 2019)

Miguel Pina e Cunha is the Fundação Amélia de Mello Professor of Leadership at Nova School of Business and Economics, Universidade Nova de Lisboa, Lisbon, Portugal. His research deals with surprising (paradox, improvisation, serendipity, zemblanity, vicious circles) and extreme (positive organizing, genocide) organizing processes. He co-authored *Positive Organizational* Behavior (Routledge, 2019), and co-edited *Management, Organizations and Contemporary Social Theory* with Stewart Clegg (Routledge, 2019).