Arius in Cyberspace: Digital Companions and the Limits of the Person

Kieron O'Hara

Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
University of Southampton
Highfield
Southampton SO17 1BJ
United Kingdom
kmo@ecs.soton.ac.uk

Version of a chapter in Yorick Wilks (ed.), Close Engagements With Artificial Companions: Key Social, Psychological, Ethical and Design Issues, *Amsterdam: John Benjamins, 2010.*

> By speech first, but far more by writing, and more again by printing, man has been able to put something of himself beyond death. In tradition and in books an integral part of the individual persists, and a part which still works and is active, for it can influence the minds and actions of other individuals in different places and at different times: a row of black marks on a page can move a man to tears, though the bones of him that wrote it are long ago crumbled to dust.
>
> Julian Huxley, *The Individual in the Animal Kingdom*, 1910

Introduction

There is a fundamental problem with the Turing Test (Turing 1950). It is designed to help decide whether a machine has genuine intelligence comparable to that of a human, and Turing asks us to imagine that it try to simulate real human performance with an interface preventing the use of anything other than performance to distinguish human and machine. It is pronounced intelligent when its performance is reliably indistinguishable from a human comparator.

This has been a profoundly influential view, especially in the field of artificial intelligence. It is not obviously impossible – the Loebner Prize is awarded annually to chatbots which can fool some human judges into thinking they are human, although none has yet managed to convince sufficiently many judges to be pronounced as having 'passed' the Turing Test itself. However, the structure of the test ignores the obvious aspect of a machine – that it is a machine and is not human.

Machines are usually created to go *beyond* human capabilities, which is why the Turing Test is an unrealistic measure of *machine* intelligence. Why would we ever want a machine to imitate human intelligence exactly? Without context, there is no real measure of success. The Loebner Prize provides a sort of context, but a rather self-referential one; the achievement of convincing 30% of the judges (or whatever artificial threshold one finds persuasive) is the end, rather than the beginning, of that particular technological route. On the other hand, a chatbot that can cope very well in a restricted context (e.g. dealing with enquiries to a bank) tends to be regarded as untintelligent because of the restriction which gives point to all its endeavours. In an

obvious analogy, a trowel that passed a gardening version of the Turing Test by being sufficiently like a human hand for digging would be a very bad trowel – what is needed is a piece of steel that can dig more efficiently and without injury. The nature of the trowel as machine of itself removes any requirement to perform like a human equivalent, and this is as true in the production of intelligent behaviour as in any other activity.

It is also incorrect, as I shall argue in this chapter, to take this to mean that the distinction between human and machine is clear and absolute. The exhibition by a machine of machine intelligence should be judged by standards that respect the contexts of design and use, standards that are likely to be very different from those we apply to human beings.[1] That does not mean, however, that it is trivial to draw distinctions between humans and the technologies we employ. Artificial cognitive support is not best understood as a separate adjunct to human cognition; the former strongly affects the nature of the latter.

The issue arises when we consider the development of *companions*, artificial agents which act as intermediaries between ourselves and a complex online world. Yorick Wilks and colleagues are exploring the science and social science of this possible technological future (Wilks 2005, 2006, Benyon & Mival 2008), and predictions of how the field will develop are bound to be premature and speculative. In this chapter I assume the success of the research programme, and will explore some of the issues that may arise if companions are routinely deployed in human societies.

When I use the term 'companion' I will not use the term in its usual English meaning, but rather refer specifically to an agent (a) with a telos determined by the intentions of an individual human being, and (b) which contains a great deal of knowledge about that individual, obtained from the individual. That individual I shall refer to as the 'original'. I will generally assume that a companion is associated with a single original, although there is no particular reason why it should not be more promiscuous and associate with a range of originals, integrating their memories and knowledge. I will consider some of the implications of this, but for simplicity I will usually assume a one-to-one association.

Companions need not be human-like or humanoid; there is nothing about companionship that demands a Turing Test sort of approach, and no reason why an artificial companion should ape the performance of human (or animal) companions. There are many human companions whose performance is lamentable even within constrained environments, such as K's ludicrous assistants Artur and Jeremias in Kafka's *The Castle* whom K drives from his service because they cannot steer him through the castle's bureaucracy. One feels that artificial companions might have been more useful.

I shall take a very optimistic view of the development of companions, to sharpen the philosophical debate. There is good reason, for example from robotics deployment, to believe that such artefacts, possibly embodied in agreeable forms such as Wilks'

[1] Of course, it is highly misleading to say that we 'apply standards' to humans to determine that they have human intelligence. It is not usually true to say that our interactions with other humans commence only when they have passed some test or exam. It only occurs to us to even think that we might have such standards when we find ourselves in highly unusual contexts when the existence of intelligence is in question – such as when we find ourselves judging the Loebner Prize. Note that the standards to which I refer here are standards to determine whether the object has a 'human' intelligence at all; there are several standards by which we judge the *level* of an already accepted human intelligence.

suggestion of a furry handbag, could be popular. As Sherry Turkle argues in the introduction to this volume, we will be challenged to see robots as "evocative objects." Wilks' idea is that a companion would be an intelligent interlocutor[2] able to process conversational language; their owners would chat to them while they build up a robust model allowing them to represent the owner in complex dealings with fast-changing technologies. Complexity and change are among the highest of the barriers to entry of the Internet and the Web, and will continue to be so for many sectors of the community. A companion is a relatively stable entity with which the technophobic can identify and bond. The companions research programme is obviously at an early stage, but it is an attractive potential solution to some of the more severe digital divides.

Autonomous agents that do one's bidding are already technological possibilities – shopbots and negotiation agents are fruitful areas of current research – but not particularly challenging to current ideas of identity and personhood. They may affect one's person – for example, they may bring on feelings of alienation or paranoia – but the current state of agent technology does not raise questions about what is 'internal' and what 'external' to the person.

However, a companion would possess a lot of information about its original from a first-person perspective. Much of its content would be unavailable from any other source. Depending on their skills of mimicry (and there is evidence that mimicry is an important support for building trust relationships – Pentland 2008), they may even sound like their original or use their characteristic speech/writing patterns. These agents would contain very rich representations of their originals, and would share much of the background information. In this chapter, I shall defend the idea that companions developed to this level of technological advancement – particularly if they were widespread throughout a society – will raise issues for our understanding of identity and personhood. The way to resolve these issues will depend on many contingent, contextual and cultural factors, and as concepts like 'cognition' expand we are never forced to apply them in a particular way, so this chapter is neither a prediction nor a strongly normative recipe. My main aim is to illustrate some interesting uncertainties in the relationship between humans, society and technology.

The structure of the chapter is as follows. First, I shall defend the idea that a full description of human psychology or cognition must deal with more than the state of a person's head/brain/mind (i.e. rejecting what has been called 'methodological individualism'). Second, I shall look at issues of identity that arise from augmentation of the individual with technology and the creation of companions. These two discussions will, I suggest, open up the possibility of some kind of internal or organic relation between a person and his or her companion. Third, I shall consider some of the possible types of relationship that might hold between a companion and its owner.

---

[2] 'Intelligent' can appear in scare quotes if the reader is sceptical about machine intelligence or keen to distinguish it from that of a human. In this case, the machine intelligence would consist, at a minimum, in abilities to (a) parse and make sense of human conversation, (b) adapt user models in the face of input information, and (c) develop courses of action that serve the original person's interests. Depending on the level of trust involved, the companion might also need to be able to manipulate the original's passwords, or even bank accounts. The main thing is that the original cannot instruct the companion at any level of detail precisely because *ex hypothesi* the user is unable or unwilling to attend to the detail of his or her online interactions. If the original could do that, he or she would not need the companion in the first place. In AI terminology, the interactions between original and companion take place at the knowledge level.

Fourth, I shall look at some of the issues of authority that we can expect to arise. The chapter ends with a brief conclusion.

Changing Boundaries

The methodological individualism suggesting that a person should be considered in isolation from his or her environment, a position perhaps held most purely by Fodor in his theory of 'methodological solipsism' (Fodor 1980), has led to a number of related ideas about psychology and technology. It is a root of recurrent difficulties in motivating a convincing rebuttal to the idea that we are merely 'brains in vats', or otherwise persistently and pathologically misled about the nature of our surroundings, a sceptical problem that took on a new lease of life when philosophers, cognitive scientists and practitioners of artificial intelligence were able to exploit analogies between hardware/body and software/mind, and convinced themselves that humans were equivalent to computers which could be, and frequently were, misled about the outside world.

The arguments for and against this position have been well rehearsed. What I want to argue in this section is that the simple analogy will appear inadequate in a world marked by wide deployment of effective companion technology. In this section I will focus on two issues that have impact in such a world: the increased tendency towards remoteness and the 'disappearance of the body' in social interaction, and the persistent habit humans have had of outsourcing or supporting mental function by suitably crafting the environment. These undermine straightforward distinctions between mind and body, human and machine, or the real and the virtual.

The Disappearance of the Body and Globalised Trust

The proportion of significant face-to-face contacts is falling all the time in what has been called by sociologists the disappearance of the body (Giddens 1990). We communicate by phone, email, letter, text; increasingly many interactions are mediated through digital technology.

Hence identification is a serious issue. In a series of face-to-face meetings, it is a trivial matter of memory to check whether someone was the same on each day; personal appearance is an effective biometric. One can be fooled by identical twins or a master of disguise, but face, voice and mannerism recognition is trivial for a human and our society has augmented human evolution with other methods to deal with less familiar persons – signatures, seals and passwords.

None of these standard high-bandwidth methods will work for the absent presence. Instead, we use technologically-constructed versions of ourselves, or *avatars*. The avatar must leave a trace behind which provides a trail back to the original person whose body has disappeared, so that credentials can be compared.

Each time a new technology appears allowing people to communicate without an immediate physical presence, a new abstraction is created – an email log, a digital representation of non-verbal communication, a certificate of trustworthiness or whatever. As fewer demands are made of our physical presence, we need increasingly many of these abstractions. As the world has adapted to the disappearance of the body, these informational avatars have become ever more prominent. One often needs a bank account or credit card to be allowed to perform certain actions, sometimes unrelated to financial transactions. Mobile phones are developing into multi-purpose trusted identifiers, and technologies such as Near Field Communication (NFC) will

boost this trend. Gradually, technological mediation will cease to be a method of securing traditional interactions, and will become a necessary precondition of existing in the modern world (e.g. collecting government benefits, or accessing one's bank account).

This process of abstraction is part of a general move, characteristic of modernity, towards expanding the extent of trust. As I have argued elsewhere (O'Hara 2004a, 75-94), trust in society, the ordinary social glue that allows potentially risky interactions from which both sides stand to gain to take place, is generally bootstrapped by a localist model of trust in personal acquaintance (no doubt kick-started by a baby's instinctive and unconditional trust of its parents – Baier 1994). Nevertheless, this local trust is heavy on resources for monitoring, and limits the extent of one's dealings to those with whom one is personally acquainted. Proxies for acquaintance can emerge, such as kin relations and reputations, but even so it has been argued (Fukuyama 1995) that societies whose trust models are local and tribe- or kin-based tend to suffer large opportunity costs and are less prosperous as a result.

Trust needs to be globalised, in order that we can trust trustworthy people with whom we are not acquainted. This is a complex matter, usually involving the intervention of institutions to certify trustworthiness and administer sanctions to wrongdoers. The disappearance of the body is a sign of globalised trust – tokens, including brands and logos, are used to associate a person (or organisation) with an interaction, thereby enabling him or her (or it) to be traced in the event of a betrayal. In the digital age abstractions are designed whose properties ensure that the interaction can be trusted by those who take part. The combination of hardware and software on a credit card, together with some personal knowledge such as a password or a PIN number, or a physical token such as a signature, fingerprint or photograph, enables a retailer to trust someone whom they have never met before.

The companion is an example of a technology for globalising trust. It would intercede 'for' its original, and represent it in some transactions. That means that it must have a strong connection with its original, and would pursue the original's interests. Donna Haraway has urged that this is a cold war model of interaction between person and machine – "biological evolution fulfils itself in the evolution of technology" (Haraway 2004b, 299) – and argues that an ethics of care for the "hybrid machine-organism" is more appropriate. Hybridism is an issue, as we have already argued in relation to the Turing Test and the spurious distinction between human and machine; we discuss it in the next subsection. However, the companion is best taken as a *component* of the hybrid, and so it makes sense to try to understand it as something whose purposes are bound up with those of its original. We can talk of the companion as having a *function*, which will in some way be bound up with the interests of the individual, and to represent the individual in a world of globalised, anonymous trust.

Extended Agency and the Export of Memory and Cognition

Such abstractions are part of a general cognitive strategy used by humans to facilitate action in the world which has helped hone basic reasoning processes using embodiedness, socialisation and the environment. The basis of human reasoning, as of all higher animals, is that of fast pattern-matching by multiple neural systems using various none-too-intuitive methods which have evolved opportunistically. Certain types of message become extremely embedded in behaviour and instinct, as cognitive structures and social patterns co-evolve. Many animals modify their environment, for example, by marking their territory with a scent; socialised animals of the same

species then become extremely good at recognising and identifying the scent. In humans face and expression recognition is particularly sensitive, as they are used for communication both conscious and unconscious. In each case, the ability to process particular aspects of reality is heightened in order to aid socialisation and communication. The work of adapting the environment can be outsourced to organisms of other species who co-evolve with us, ranging from intestinal flora to highly socialised animals such as dogs or horses (Haraway 2004b).

Andy Clark has argued that the hallmark of specifically human cognition is not found in fundamentally neural processes, but rather in "our amazing capacities to create and maintain a variety of special external structures (symbolic and social-institutional). These external structures function so as to complement our individual cognitive profiles and to diffuse human reason across wider and wider social and physical networks whose collective computations exhibit their own special dynamics and properties" (Clark 1997, 179). Augmenting theories of extended cognition can result in complex accounts of extended agency and extended (moral) responsibility (Hanson 2009).

Substantial problem-solving work is offloaded to the environment, which is shaped by human action (including interpreting some objects as symbols) in order to act as either a corporate memory or as a short cut in the problem-solving process. This shaping/restriction of context substantially reduces cognitive load (O'Hara & Shadbolt 1997); the more aspects of the environment that can be assumed, the less information processing needs to be done, because facts about the environment can be asserted rather than derived. To use Clark's term, reasoning is *diffused* across the environment by human action, allowing us to succeed by using our intelligence more intelligently. An alternative way of looking at this is to see the mind as the human brain plus the engineered environment; the engineering of the environment in effect allows us to include some reasoning as common property.

When the engineered environment includes virtual representations, the physical, the digital and the cognitive become harder to separate out, as for example with military representations of battlefields presented as augmented perceptions to soldiers. One ambitious gadget is a helmet whose visor provides a transparent display mapped onto the view through it, showing street and landmark names, identifying what is out of the line of sight and identifying threats such as sniper positions established by reconnaissance (Bulman et al 2006, 119-122). In such an environment, a central example of what Turkle in the introduction to this volume calls "our culture of simulation", it would be just as much a mistake to postulate a strict separation of external reality and cyberspace as it would be to assume that real life was just another window on the visor (Žižek 1997, 132). The two 'realities' are intertwined, linked by the cognition of the soldier.

Theories of individualistic rational mind are hard to justify in the abstract precisely because of the complexity of inferring everything from scratch about the context, but when choices are severely constrained they look much more plausible (Satz & Ferejohn 1994). Reasoning, on this reading, becomes a communal activity, even when the locus of activity appears to be the individual mind. As David Hume reminds us, all human action depends absolutely crucially on *inter*action with others, whether we acknowledge it or not.

> The mutual dependence of men is so great, in all societies, that scarce any human action is entirely compleat in itself, or is performed without some

> reference to the actions of others, which are requisite to make it answer fully the intention of the agent. The poorest artificer, who labours alone, expects at least the protection of the magistrate, to ensure him the enjoyment of the fruits of his labour. He also expects, that, when he carries his goods to market, and offers them at a reasonable price, he shall find purchasers; and shall be able, by the money he acquires, to engage others to supply him with those commodities, which are requisite for his subsistence. In proportion as men extend their dealings, and render their intercourse with others more complicated, they always comprehend, in their schemes of life, a greater variety of voluntary actions, which they expect, from the proper motives, to co-operate with their own. … A manufacturer reckons upon the labour of his servants, for the execution of any work, as much as upon the tools, which he employs, and would be equally surprized, were his expectations disappointed. In short, this experimental inference and reasoning concerning the actions of others enters so much into human life, that no man, while awake, is ever a moment without employing it. (Hume 2007, 64-5)

Trust in the reliability of others makes human society possible (Seabright 2004), and indeed without it it is hard to see how any kind of moral relationships could even be definable (Baier 1997). That reliability extends to interpretations of functional objects – for instance, money, a vital store of value without which restructuring of the environment would remain rudimentary, and a long way short of the sophistication of modern urban environments where so much thinking is, as it were, done for us in concrete.

The symbolic significance of our alteration of the environment varies from the basic (e.g. flattening floors to make walking easier) to highly complex intermediaries such as money or language. Digital technology gives an extra dimension to cognitive scaffolding – it allows the creation of artefacts, such as Wilks' companions, which can act independently of human control. Of course, the actions of artificial agents require a human-created and mediated environment (as do human actions), but they do not need *direct* command.

Conversation and Independence

Independence does not need digital technology to support it. As Julian Huxley reminded us at the beginning of this chapter, all sorts of communicational methods have been used to socialise or outsource mental function. Most of these have been fiercely resisted, because of fears of a decline in that mental function, although it often turns out that other functions can be enhanced given the lightened cognitive load. The dilemma was put in a particularly clear form by Plato, through the mouth of Socrates, in the *Phaedrus*.

> You know, Phaedrus, writing shares a strange feature with painting. The offsprings of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words. You'd think they were speaking as if they had some understanding, but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever. When it has once been written down, every discourse roams about everywhere, reaching indiscriminately those with understanding no less than those who have no business with it, and it doesn't know to whom it should speak and to whom it should not. And when it is faulted and attacked unfairly, it always needs its

> father's support; alone it can neither defend itself nor come to its own support. (Plato 1997a, 552 [275de])

This is a fascinating and ironic passage. Socrates, of course, did not write anything down and may well have actually held views not unlike these. On the other hand, his immortality is guaranteed by written accounts of him by Plato, Aristotle, Xenophon and Aristophanes. His point in the *Phaedrus* applies just not only to relatively primitive technology such as writing, but also to more involved digital technologies such as companions. His reservations can be summarised as the following points.

1. Writings cannot explain themselves.
2. The circulation of writings cannot be restricted.
3. Writings cannot defend themselves against attack.

Successful written communication is grounded by oral communication, argues Plato (a written response to a written attack would be recursively vulnerable to his original argument once more). For these three enumerated reasons, an author loses control of his or her output. Only through engaged conversation does wisdom follow. Writing can only "remind those who already know what the writing is about." It has no independent life. More abstractly, Plato grounds interaction in physical presence, and is worried by absence (such as the disappearance of the body described at the beginning of this section).

As Havelock pointed out (Havelock 1963), Plato, presumably like the historical Socrates, was wrestling with the opportunities and threats of a new literate world. Other types of cognitive scaffolding as described by Clark have similar properties to writing, and are also unresponsive to physical presence. From Clark's perspective, this is not a serious problem – there is no requirement for a cognitive support to "know to whom it should speak", and no reason why it should do more than remain "solemnly silent." It need not be able to act independently.

New technologies are beginning to change this landscape (O'Hara 2004b). In particular, technologies that can exploit the metadata attached to information can be more responsive by managing permissions and allowing reasoning about policies to protect privacy etc (Weitzner et al 2005), by gathering or collecting supporting information (Blythe & Gil 2004), or by tailoring publishing strategies to the particular requirements of the reader (Weal et al 2007). The reader can also bring his or her own ideas and intellectual structures to the interaction, using ontologies to mediate and possibly to uncover content in the discourse that was unknown to the writer (Fensel 2004).

The Linked Data Web, more commonly if less accurately known as the Semantic Web (Shadbolt et al 2006), might supply this revolution single-handed if it takes off as hoped, but even if not, personalising technologies could allow surfing the Web to seem more like a conversation than hitherto, on an *ad hoc* basis. The movement loosely known as Web 2.0 has shifted the balance away from writers and towards readers, with the blogosphere already looking like a conversation, and with aggregation techniques allowing us to see the 'wisdom of crowds' (Surowiecki 2004) giving concrete form to a General Will of the type that Rousseau described, very separate from the wills, wisdom or intelligence of its components. There is not a great deal of automation in the Web 2.0 space, but the difference from what we might call the 'traditional' Web (if something can be both traditional and under two decades old) is clear.

As such technologies become more effective and commonplace there is clearly a place for companions, assuming that the technology they require continues to develop. The lack of physical presence of a human being is already, as I have argued, less of a handicap than Plato anticipated, and many of the shortcomings of the disappearance of the body are being addressed by technology. Artefacts can defend themselves against attack, can restrict their audiences, and can explain themselves. Companion technology is part of a general movement in this direction.

Summary

To summarise this section, we have argued that traditional distinctions are being broken down by social and technological development. Brain, body and mind are strongly linked by causal ties, and it is a characteristic of human intelligence to alter the environment – which includes intellectual and virtual environments as well as the physical world – to support cognition and provide intellectual resources. Mind is embodied, but not just in the brain. Furthermore, these resources, being reusable, link minds and persons. Social development, in the form of increased and more widespread trust, is also facilitated.

It may well not be long before stores of digital information begin to demonstrate independence of the minds that created them, not only in terms of their content (being assembled by a number of people and thereby transcending the knowledge or memory of each individual), but also in terms of action. In that case, the extended agency position as sketched for example by Clark would be significantly complicated. In the next section, we will consider some of the implications.

Augmentation and Companionship

Centralisation of intelligence has been a common assumption throughout the prehistory and history of cognitive science. Most works in the early philosophy of artificial intelligence assumed that a single controlling executive would be required in order to produce intelligent behaviour artificially – the so-called 'good old-fashioned AI' or GOFAI programme whose individual successes never added up to the achievement of the goal. The idea of distributed intelligence came later, and also provided some success, but the general assumption that artificial intelligence would be realised in standalone devices remained until the World Wide Web came along. Even then the successes of statistical techniques such as PageRank surprised many of the major figures in the field. The Semantic Web still struggles to evade the accusation of adopting the same fallacies as GOFAI (Spärck Jones 2004, Wilks & Brewster 2006), although more recent discussions of the SW emphasise the importance of local ontologies, small projects and avoiding overweening ambition (Berners-Lee et al 2006, Shadbolt et al 2006, Alani et al 2008).

Meanwhile, other types of web presence have mutated from individuals to less well-defined collectives. The lifelog, a comprehensive and indiscriminate record of someone's existence, has typically been seen as being strongly associated with a particular person but the arrival of social networking has resulted in lifelogs with distributed properties (O'Hara et al 2009). These developments have mirrored those in psychology which dethroned central executive function (Libet 1985, Sheth et al 2009).

Even commentators who are sensitive to technology's transgressive tendencies can find it difficult to break out of this type of methodological individualism. For instance,

Haraway in her work on cyborg culture argues that "the dichotomies between mind and body, animal and human, organism and machine, public and private, nature and culture, men and women, primitive and civilized are all in question ideologically" (Haraway 2004a, 22), and that in particular the organism/machine distinction is particularly "leaky" (Haraway 2004a, 10-11). She writes at length about the possibilities that unfold for liberating human identity when machine prosthetics become available, and about her "dream not of a common [and therefore totalitarian] language, but of a powerful infidel heteroglossia" (Haraway 2004a, 39). Nevertheless, she still places her self, and those of others to be liberated, on the same ontological and hierarchical level as the cyborgs whose imperatives and Cold War myths she wishes to rewrite.

Similarly, her work on companions focuses on dogs, which "are not a projection, nor the realization of an intention, nor the telos of anything" (Haraway 2004b, 300). This is of course an important point to make about dogs (and other companion animals, including humans), but although she eloquently unpicks the ethical imperatives that follow from the co-evolution of dogs and people, she nowhere follows the thought that in the realm of the artificial, telos *is* an important factor. In a world where Wilks-style companions are commonplace, the ethics of interacting with teleologically-specified agents may be more complex. We should not be surprised if companions evolved from standalone applications to more complex entities with complex functions; in this section I will sketch some potential implications.

Extensions in Time

As Wilks has often emphasised, the companion, being artificial, may well outlive the original person from whom it gained its content. There is no reason why in such circumstances the companion could not continue to perform its function in society as an intermediary for some aspect of the digital world. It may well even be that over time the number of such companions multiplies to create a substantial population, as for example in Charles Stross' novel *Saturn's Children*, in which the dwindling human race has created a whole range of types of robot designed to do its bidding, from labouring drones to concubine sex machines. After humans finally die out, the robots carry on following their programmed imperatives, but with enough intelligence to continue within the skeletal framework of the human society, forever trying to please humans who are now mythical ("the sad fact is, human civilization did not even break for lunch when humankind died out.").

A sufficiently intelligent companion could function as an important memorial or source of information about the original, now deceased, person. Great-grandchildren could question the companions of their great-grandparents in order to build up their knowledge of barely-remembered figures. If the companion was a repository of the preferences of the original person (for example in legal matters), then it could infer the original's likely wishes in complex cases involving disputed wills, the management of trusts, or legal documents whose measures were now out of date, for example, assuming the companion's knowledge base was reliable and properly certified as containing an accurate reflection of the original's desires. Rather than search for a living will, a doctor might consult a companion to try to determine a preference as to whether an apparently brain-dead person would wish her life support systems to be switched off. In the event that the support systems remained switched on, the companion might continue to act for her.

Such a companion would enjoy a special relationship to the original person. Important projects could continue, and the wisdom of older generations preserved and consulted. That is not to underestimate the difficulties of providing what C.P. Snow memorably called "that singular necrophilic confidence … about what a dead relative would have 'liked,'" but rather to expect a sufficiently intelligent agent with a sufficiently good model of the original person in a sufficiently well-specified and circumscribed context to be reliable, and to be certified to be so.

Federations of Identity/Identities

At least since Locke, modern philosophers have used the body as an anchor for identity, because – it so happens – it gives spatiotemporal continuity, hosts the sensory organs and the brain, and also provides the limbs that are used initially if not exclusively to investigate, explore and map the world. There is a debate as to whether the *concept* of personal identity requires the body (often carried out using thought experiments about interrupting spatiotemporal continuity or transferring someone's brain into a different body, for example), but it is at a minimum regarded as a useful proxy for the locus of identity. Nevertheless, when cognition or memory are exported to the environment, deliberate damage to that environment may be morally more significant than mere vandalism, and looks more like harm to a person. One does not have to hold a Marxian view of the relation between labour and identity to feel that the theft of a laptop computer with several years' work stored on it is somewhat more like an attack on the person than, say, the theft of an equally valuable DVD player or a more valuable piece of jewellery (Clark 1997, 215).

Identity has always been an issue in cyberspace, captured by the immortal *New Yorker* cartoon whose caption read "on the Internet, nobody knows you're a dog." Similarly, nobody knows you are a companion. Even if we don't go as far as those, such as Turkle (1995), who argue that the Internet gives us unequivocal examples of the 'decentred subject,' the case of the companion is still difficult to describe; it acts for, but is independent of, an original. Orders could be countermanded, but equally the companion does have the special relationship to the original as postulated in the previous section. There would be no reason in principle why the companion could not generate and sign legal or financial documents (although in practice of course society may refuse to let it happen), thereby even going beyond the laptop as an extension to the person. Machines already make decisions of great moral import, for example in the medical domain (Tuffs 1996).

A companion would have an important connection to its original. Even given the original person was consciously aware of its artificial nature, it would still be a significant relationship. Ties of kinship, intimacy and similarity tend to undermine or undercut more formal ties, as for example in Joseph Conrad's short story 'The Secret Sharer', where the captain of a ship shelters a fleeing murderer for whom he feels the affinity of similarity, and as a result neglects his duties and risks his ship. This basic relation can be complicated, because different companions of the same original could be merged in various ways, with the potential for sideways growth of identities, again as envisaged by Stross in *Saturn's Children* ("I haven't worn her long enough to receive more than a perfumed hint of her presence in my head," as one robotic character comments).

Such relationships would be relatively straightforward compared with a merge of companions from two different originals, or in the simpler case where a companion's data is deliberately falsified for evil purposes. Falsification of data may even be

voluntary – in the context of lifelogging, Dodge & Kitchin (2007) suggest that privacy can be preserved for a lifelogger if he or she deliberately falsifies a small proportion of the lifelog, so that no-one could be sure that information extracted from the lifelog was veridical.

The companion could be equipped with inferential capacities that take the expressed preferences of the original and develop them in unanticipated directions. It may treat the original's preferences more seriously and with more consistency (for instance, it might be more protective of the original's privacy than the original himself). It could be equipped to take information from the context or environment to amend information (for instance, 'correcting' memories that it discovers are false, augmenting its knowledge base via Google or interacting and sharing information with other companions). It could use the original person's social network to make itself into a more communal resource.

In short, the companion need not be restricted to a mirror of the original. Context, online information resources, pre-programmed information, other companions and other members of the original's social network could also help inform its knowledge base (as well, of course, as its own experience). The relationship between the identities of the companion and the original need not be simple or straightforward.

Arius 2.0

Arius was a Berber priest who died in AD336, who thought deeply about the relationship between the three manifestations of God – God the Father, Jesus the Son, and the Holy Spirit. Arianism was deeply controversial, disputed for many years in the Late Roman Empire, finally being denounced as heretical.

Arius' problem was a simple one – how to understand the Trinity. The now-official doctrine of most large Christian denominations which won the ancient theological and political argument was that God is a unity consisting of three persons, and of course this is something of a mystery for worshippers. Arius' idea was a simpler one, which is perhaps why it appealed to barbarian converts from paganism: God is three persons. God was not the Father before He begot the Son. Jesus, though divine, was not of the same substance as God, was a created being, and may even be inferior to God.

These weighty issues are somewhat beyond the scope of this chapter, but like Arius of old, if the companions idea is as successful as its advocates hope, we will be faced with a similar knotty problem of identity. Instead of the Holy Trinity, we will have to grapple with a rather more Earthbound set of *body*, *avatar* and *furry handbag*.

The body of the original person is the aspect of the new trinity with which we are familiar, the provider of spatiotemporal continuity and the Lockean ground of identity. The virtual companion is an avatar. Thirdly, there will be bodily incarnations of an avatar, whose nature would of course strongly affect the sorts of information it was able to extract from a person. Wilks playfully imagines a furry handbag to which a person chatters all day. Wilks' own Companions project[3] has experimented with both embodied Nabaztag rabbits and virtual forms, including an animated version of the British comedian Ken Dodd.[4] There is some evidence that elderly Japanese people are quite happy to interact with anthropomorphic robots, and indeed often prefer them

[3] http://www.companions-project.org/.
[4] Disconcertingly, for the British viewer of a certain age, without Doddy's distinctive Liverpool accent: http://www.youtube.com/watch?v=MUe9dUccVh8.

to non-Japanese people who cannot be programmed to avoid behavioural taboos. It has been suggested (Chrisley 2003) that the Kismet robot developed at M.I.T.[5] is so compelling as an interlocutor because it has extremely expressive eyebrows. The form the 'furry handbag' takes will be very important for the type of information that can be brought out of the human and used to populate the companion's memory.

Arius' solution was to treat each aspect of the Trinity as different and separable; the victorious Athanasian Creed treats the Trinity as three aspects of one person. Neither solution solves the issues with respect to companions. The independence of the companion makes it hard to regard it as the *same* as the original person, or merely a tool used by that person, while its strong links with the original make a story of *separation* implausible too.

Types of Relationship

It is usual to see human beings as taking a clearly identifiable place within a hierarchy or partial order of extended or augmented agency. An extended agent contains people as proper components. Hanson (2009) gives the example of two people colluding in a murder using a car, and argues that moral responsibility falls to the extended agent two-people-and-car. Though perhaps few would find this reasoning convincing, the point is that he views the individuals as components of the more complex extended agent. Even those less Catholic about what counts as an extended agent take an analogous view; a cyborg as described by Haraway, or a scaffolded mind described by Clark are basically human agents with reference to whom cognitive enhancements (whether machinery or amended aspects of the environment) are identified. Although Clark insists that a description of a person's mind requires a description of the environment, he does not directly address the question of federation of identity. Though questions like "where does the user end and the tool begin?" are "a delicate call", he does not dwell on the implications of the thought that "much of what we commonly identify as our mental capacities make … turn out to be properties of the wider, environmentally extended systems of which human brains are just one (important) part" (Clark 1997, 214-5). Are there any grounds to keep the individual human brains distinct? Does Clark's citation of human language as a tool bring other minds into important contact with our own (possibly exploiting Wittgenstein's argument that language is essentially public)? Similarly Haraway's companions are teleologically separate, agents in their own right.

Of course, the furry handbag itself (whatever form it takes) is likely to be physically separate from the body of the person,[6] but the world-involving nature of human cognition, and the human-involving nature of artificial cognition together mean that the hierarchy of companionate agency might well be much more complex than these examples suggest. The companion need not simply be an autobiographical (= solely the responsibility of the original person) recreation of the original. As noted above, the companion's knowledge base might go beyond the information provided by the original; this indeed is likely, as one of the main suggested ideas behind companions is to act as an intermediary between non-techy persons and the complex online world. In any case, it is hard for anyone to control the picture they present of themselves. For instance, Dr Johnson provided much solid input to Boswell's depiction of him, and

[5] http://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html.

[6] Likely but not necessarily. The technology required to go beyond the furry handbag to produce a companion physically linked to the original is perhaps more like Haraway's conception of a cyborg.

exercised a good deal of editorial control over the process (as Boswell admits in his *Journal of a Tour to the Hebrides*), but even so Boswell determined the final product, distilled from conversations with Johnson and others, testimony from others who knew Johnson and objective sources of information where possible. The result is somewhat mythical, and certainly not what anyone would call an objective account, but equally not one that Johnson would have endorsed despite his level of input (he did not like being called 'Dr Johnson', for example).

As Haraway argues, in general companions are separate entities in their own right, but this is not true of the mechanistic companions discussed in this chapter which derive important aspects of their meaning from their originals. We may ultimately have to think deeply about the ethical treatment of companions (as Sherry Turkle suggests in the introduction to this volume), but at least some aspects of their telos will be strongly, intimately and intrinsically related to its original. The considerations of the previous paragraph cannot be used as arguments for completely detaching the agency of the companion from the agency of the original.

There are various examples in literature and philosophy of such complex relationships between original and creation, some already mentioned. In Stross' *Saturn's Children*, 'families' of robots are able to share memories, experience and capacities, which though it does not reduce their agency to a single entity, equally it is sometimes threatening to their individuality. Plato's representations of his mentor Socrates seem to have evolved from a lifelike portrayal of the sceptical gadfly and social irritant, with a resemblance to Aristophanes' more waspish picture in *The Clouds*, to a more independent figure; first applying Socratic methods to philosophical problems without coming to direct conclusions, then taking a stronger, less sceptical stand on matters such as justice (e.g. in *The Republic*), and ending up finally as a mouthpiece for Plato's mature and complex philosophy.[7] The later dialogues use reasoning methods with which Socrates might have been sympathetic, but which appear to have developed after his death (such as the inductive method of studying the properties of objects in order to discover what they have in common), and one can't help thinking that the real Socrates would have made great sceptical play with some of Plato's mature opinions.

Nevertheless, the written Socrates is importantly constrained by the original – to the extent, as we have seen, of paradoxically arguing against writing as a means of preserving philosophical argument. The newly literate world in which Socrates and Plato found themselves was as challenging to its inhabitants as the digital world is to us (Havelock 1963), and Plato's use of Socrates as a long-running character, living on for maybe half a century after taking hemlock, must have been a literary and philosophical conceit of impressive originality. In the *Second Letter*, the author, supposedly Plato, tells Dionysius that "there is no writing of Plato's, nor will there ever be; those that are now called so come from an idealized and youthful[8] Socrates" (Plato 1997b, 1639 [314c]). Most scholars believe that Plato's *Second Letter* is a forgery, but the point still stands that a near-contemporary was prepared to impute agency to the Platonic Socrates, however metaphorically.

Some of the more ambitious twentieth century attempts to explore the limits of character and fiction also provide fertile ground. Norman Mailer's *Ancient Evenings* circles around a series of reincarnated characters called Menenhetet, while Mailer

[7] There are many uncertainties here, but this general statement sums up the rough consensus.
[8] Or, alternatively, 'modernized'.

tried, and failed, to produce an eight-volume series of novels on social and philosophical themes all featuring a romancing fake Irish adventurer and observer called Sergius, or sometimes Cassius, O'Shaugnessy.[9] Malcolm Lowry also played with a series of interlinked and recognisably similar lead characters in a series of novels (only one of which was completed at the time of his death) portraying a spiritual journey from Hell through Purgatory to Heaven. These characters were lightly fictionalised aspects of Lowry himself, and their misadventures closely mirrored his own. They are of course fictional, but independent critical treatment of any of them would be inappropriate given the nature of Lowry's *oeuvre*. In these cases, the fictional characters can even be seen without too much hyperbole as 'companions' in Wilks' sense, animated by prose, who pilot their authors through philosophical, political and spiritual malaises.

The relationship between a person and his or her companions is a complex one to describe. When the companion is adding substantially to the original's cognition, carrying out important cognitive or social functions, then it is extremely plausible to see it as making a contribution to the identity or the personhood of the original, as would be argued by Clark or Hanson, for instance. When the companion becomes more intelligent, capable and autonomous, then it might also be seen at the same time as having an identity or personhood (however derivative, metaphorical or incomplete) of its own. The simple hierarchies that are usually taken to characterise relations of identity between people and artificial agents become harder to sustain, and the relationship between companion and original will be more tangled, interwoven and context dependent.

Authoritative Representations/Representatives

The aim of the companion is to perform some task(s) for the original person, which it will do using inbuilt designed capabilities informed by knowledge absorbed from the original person via knowledge acquisition procedures and conversational extraction techniques. Such tasks might include negotiating online prices of goods, organising photographic records, constructing biographies to order, and so on. To perform those tasks in the stead of the original, the companion must have some kind of authorisation which will presumably usually be provided directly by the original person using a secure certificate, but this will not always be necessary or (for example in the case of the use of a companion to sort out a posthumous legal tangle) possible. The context in which the companion is used to represent a person will determine whether it has the ability properly so to do.

Assuming that the use of a companion in a context is acceptable, the key factor then is whether its model or representation of the original person is *sufficiently accurate*. In the offline world, the usefulness or acceptability of a representative, such as a solicitor, will depend crucially on how well he is briefed about his client, and an analogous issue arises for the companion. Is the model that the companion holds 'truthful'?

[9] One novel (*The Deer Park*) and a couple of short stories remain of this scheme. One reason for the failure of Mailer's Balzacian project was that he gradually morphed into this idealised observer figure himself, complete with fake Irishisms, with the projected novels replaced by long pieces of journalism in the first person – an indication of the closeness of the literary 'companions' and the original in this case.

The problem is that, where personality and opinions are concerned, the notions of accuracy and truthfulness are hard to pin down. The methods used by the companion for knowledge acquisition and modelling will certainly be a factor. The emphasis on methodology is not unusual – for example, in psychoanalysis, the methods for understanding the patient's subconscious are assumed to be superior to the patient's own assessment of his situation (the subconscious is deemed to be more truthful than the conscious personality).

In the virtual world, consciously-constructed avatars can be seen as more authoritative than the public face of the original person, even by the original herself. One can, in constructing an avatar, enjoy the distance and the relative lack of risk by exaggerating one's personality, pretending to be brave or beautiful or psychopathic or ruthless, or alternatively, one can construct a persona that is 'more' oneself than the compromised self of everyday existence. One can express aspects of oneself that one would normally bury in order to keep one's complex offline social life together. These two ways of relating to one's avatar are intertwined (Žižek 1997, 137-8). Inhibition and shame are (usually) absent from online existence, and one can give voice to the hidden aspects of one's personality as long as one is aware of the lack of consequence. But, as Žižek argues, it would be a mistake to focus on the 'truth' about these matters.

> When a man who, in his real life social contacts, is quiet and bashful, adopts an angry, aggressive persona in virtual reality, one can say that he thereby expresses the repressed side of himself, a publicly non-acknowledged aspect of his 'true personality' – that his 'electronic id is here given wing'; however, one can also claim that he is a weak subject fantasizing about more aggressive behaviour in order to avoid confronting his real life weakness and cowardice. (Žižek 1997, 137-8)

As Erving Goffman argued decades ago (1959), our negotiation of everyday life is akin to playing a series of roles as an actor, and one needs a backstage to express certain aspects of one's personality that cannot be shown in public. For instance, in a working context:

> The backstage language consists of reciprocal first-naming, co-operative decision-making, profanity, open sexual remarks, elaborate griping, smoking, rough informal dress, "sloppy" sitting and standing posture, use of dialect or sub-standard speech, mumbling and shouting, playful aggression and "kidding," inconsiderateness for the other in minor but potentially symbolic acts, minor physical self-involvements such as humming, whistling, chewing, nibbling, belching and flatulence. (Goffman 1959, 128).

Of course, this is only backstage with respect to the interaction with the public or with 'the boss'; this 'backstage' is 'onstage' with respect to colleagues, and for that interaction another 'backstage' is needed (perhaps the family, or friends in the pub). This social management of emotions has been taken online with virtual reality acting as yet another 'backstage' resource.

Žižek writes about constructed avatars which the original person both constructs and 'acts through', a personality which the person adopts, for example in an online game such as *Second Life* or *World of Warcraft*, but the companion takes the externalisation further by automating behaviour and ultimately acting autonomously. The original person plays no part in the actions of the companion, though he may help 'steer' it

through interaction, conversation and so on. This distancing sharpens the question of authority, or, as Turkle puts it in the introduction to this volume, the "notion of authenticity" becomes "threat and obsession, taboo and fascination". The culture of simulation is tested at its limits; the question is not, as in the Loebner Prize, how to fool a group of judges in a series of artificial tests, but rather how to operate successfully in society. There is no Turing Test requirement to masquerade as a 'genuine' intelligent being – the companion need not and should not try to conceal what is most germane about itself, that it has particular capacities programmed into it. Indeed, if it were embodied in the form of a furry handbag, it would be very unlikely to succeed.

The question instead is whether the companion can actually perform genuine functions in a complex, dynamic social world, consistent with its teleological requirement to represent or intermediate for the original person. That will require a set of person-centred knowledge acquisition methods, and set of world-centred procedures and knowledge bases. Criteria for success will vary, but the closer the companion comes to making a genuine contribution to our personhood, the more pressing this question will be. The companion will need a model of the original person that 'fits' very well, and to be able to deploy that model in a complex world. This notion of 'fit' between model and original is the crux of the matter; the better the fit, the wider the companion's legitimate and authoritative deployment can be, and the wider the legitimate deployment, the more influence on the original's personhood it is likely to have.

Conclusion

The relationship between a companion and a person will become increasingly problematic as companion technology improves and as models of users become increasingly sophisticated, and the simple dichotomies that make the Turing Test so plausible as a means of determining intelligence will become harder to maintain. As with any kind of content-storing technology, such as writing, or in more recent years laptops, the amount and quality of cognition that a human can 'export' to these outside technologies is significant. The 'person' or 'agent' can be seen as an extended system including the technologies as well as the human, in which the technologies, among other things, can help in the extension of trust towards the human. If companion technology becomes so sophisticated that the companion is capable of acting independently in its original's interests, then the relationship becomes even more difficult to describe, because there will be a case for regarding the *companion* as a person or an agent too. Much will depend on the context, circumstances and assumptions made at the time, particularly with regard to the trust relations – as Wittgenstein reminds us, there is nothing in psychological concepts such as 'person' that tells us how we should go on in new circumstances or language games. In this chapter, I have tried to produce a preliminary sketch of some of the considerations that might be brought in, and to indicate appropriate background theories that will illuminate the new situation. Nevertheless, the main determining factors are the technologies themselves, and the social practices within which they function.

Acknowledgements

The work reported in this chapter was partly supported by the projects LiveMemories – Active Digital Memories of Collective Life, Bando Grandi Progetti 2006, Provincia

Autonoma di Trento, and the EU FET project Living Knowledge, (http://livingknowledge-project.eu/), contract no. 231126.

References

Harith Alani, Wendy Hall, Kieron O'Hara, Nigel Shadbolt, Martin Szomszor & Peter Chandler (2008). 'Building a pragmatic Semantic Web', *IEEE Intelligent Systems*, 23(3), 61-68.

Annette C. Baier (1994). 'Trust and antitrust', in *Moral Prejudices: Essays on Ethics*, Cambridge MA: Harvard University Press, 95-129.

Annette C. Baier (1997). *The Commons of the Mind*, Chicago: Open Court.

David Benyon & Oli Mival (2008). 'Scenarios for companions', Austrian Artificial Intelligence Workshop, Vienna, http://www.companions-project.org/downloads/CompanionsScenarios.pdf.

Tim Berners-Lee, Wendy Hall, James A. Hendler, Kieron O'Hara, Nigel Shadbolt & Daniel J. Weitzner (2006). 'A framework for Web Science', *Foundations and Trends in Web Science*, 1, 1-130.

Jim Blythe & Yolanda Gil (2004). 'Incremental formalization of document annotations through ontology-based paraphrasing', 13th International World Wide Web Conference, New York, http://www.isi.edu/~blythe/papers/pdf/www04.pdf.

J. Bulman, B. Crabtree, A. Gower, A. Oldroyd & J. Sutton (2006). 'Mixed-reality applications in urban environments', in Alan Steventon & Steve Wright (eds.), *Intelligent Spaces: The Application of Pervasive ICT*, London: Springer-Verlag, 109-124.

Ron Chrisley (2003). 'Embodied artificial intelligence', *Artificial Intelligence*, 149, 131-150.

Andy Clark (1997). *Being There: Putting Brain, Body and World Together Again*, Cambridge MA: Massachusetts Institute of Technology Press.

Martin Dodge & Rob Kitchin (2007). 'Outlines of a world coming into existence: pervasive computing and the ethics of forgetting', *Environment Planning B, Planning and Design*, 34, 431–45.

Dieter Fensel (2004). *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, 2nd edition, Berlin: Springer-Verlag.

Jerry Fodor (1980). 'Methodological solipsism considered as a research strategy in cognitive psychology', *Behavioral and Brain Sciences*, 3, 63-73.

Francis Fukuyama (1995). *Trust: The Social Virtues and the Creation of Prosperity*, New York: Basic Books.

Anthony Giddens (1990). *The Consequences of Modernity*, Cambridge: Polity Press.

Erving Goffman (1959). *The Presentation of Self in Everyday Life*, Garden City, NY: Doubleday Anchor.

F. Allen Hanson (2009). 'Beyond the skin bag: on the moral responsibility of extended agencies', *Ethics and Information Technology*, 11, 91-99.

Donna Haraway (2004a). 'A manifesto for cyborgs: science, technology, and socialist feminism in the 1980s', in *The Haraway Reader*, New York: Routledge, 7-45.

Donna Haraway (2004b). 'Cyborgs to companion species: reconfiguring kinship in technoscience', in *The Haraway Reader*, New York: Routledge, 295-320.

Eric A. Havelock (1963). *Preface to Plato*, Oxford: Basil Blackwell.

David Hume (2007). *An Enquiry Concerning Human Understanding*, Oxford: Oxford University Press.

Benjamin Libet (1985). 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behavioral and Brain Sciences*, 8, 529-566.

Kieron O'Hara (2004a). *Trust: From Socrates to Spin*, Duxford, Icon Books.

Kieron O'Hara (2004b). 'Socrates, trust and the Internet', 2nd International Conference on Speech, Writing and Context, Kansaigaidai, Japan, http://eprints.ecs.soton.ac.uk/15836/.

Kieron O'Hara & Nigel Shadbolt (1997). 'Interpreting generic structures: expert systems, expertise and context', in Paul J. Feltovich, Kenneth M. Ford & Robert R. Hoffman (eds.), *Expertise in Context: Human and Machine*, Menlo Park CA/Cambridge MA: AAAI Press/Massachusetts Institute of Technology Press, 449-472.

Kieron O'Hara, Mischa M. Tuffield & Nigel Shadbolt (2009). 'Lifelogging: privacy and empowerment with memories for life', *Identity in the Information Society*, 1.

Alex Pentland (2008). *Honest Signals: How They Shape Our World*, Cambridge, MA: Massachusetts Institute of Technology Press.

Plato (1997a). 'Phaedrus', in John M. Cooper (ed.), *Plato: Complete Works*, Indianapolis: Hackett Publishing Company, 506-556.

Plato (1997b). 'Second letter', in John M. Cooper (ed.), *Plato: Complete Works*, Indianapolis: Hackett Publishing Company, 1636-1640.

D. Satz & J. Ferejohn (1994). 'Rational choice and social theory', *Journal of Philosophy*, 91, 71-87.

Paul Seabright (2004). *The Company of Strangers: A Natural History of Economic Life*, Princeton: Princeton University Press.

Nigel Shadbolt, Tim Berners-Lee & Wendy Hall (2006) 'The Semantic Web revisited', *IEEE Intelligent Systems*, 21(3), 96-101.

Bhavin R. Sheth, Simone Sandkühler & Joydeep Bhattacharya (2009). 'Posterior beta and anterior gamma oscillations predict cognitive insight', *Journal of Cognitive Neuroscience*, 21, 1269-1279.

Karen Spärck Jones (2004). ' What's new about the Semantic Web? Some questions', *SIGIR forum*, 38, http://www.sigir.org/forum/2004D/sparck_jones_sigirforum_2004d.pdf.

James Surowiecki (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*, London: Little, Brown.

A. Tuffs (1996). 'Eurotransplant to allocate kidneys by computer', *The Lancet*, 347(9011), 1326.

Alan M. Turing (1950). 'Computing machinery and intelligence', *Mind*, 59, 433-460.

Sherry Turkle (1995). *Life on the Screen: Identity in the Age of the Internet*, New York: Simon & Schuster.

Mark J. Weal, Harith Alani, Sanghee Kim, Paul H. Lewis, David E. Millard, Patrick A.S. Sinclair, David C. De Roure & Nigel R. Shadbolt (2007). 'Ontologies as facilitators for repurposing Web documents', *International Journal of Human-Computer Studies*, 65, 537-562.

Daniel J. Weitzner, Jim Hendler, Tim Berners-Lee & Dan Connolly (2005). 'Creating a policy-aware Web: discretionary, rule-based access for the World Wide Web', in E. Ferrari & B. Thuraisingham (eds.), *Web and Information Security*, Hershey PA: Idea Group Inc, http://www.w3.org/2004/09/Policy-Aware-Web-acl.pdf.

Yorick Wilks (2005). 'Artificial companions', *Interdisciplinary Science Reviews*, 30, 145-152.

Yorick Wilks (2006). *Artificial Companions as a New Kind of Interface to the Future Internet*, Oxford Internet Institute Research Report 13, http://www.oii.ox.ac.uk/research/publications/RR13.pdf.

Yorick Wilks & Christopher Brewster (2006). 'Natural Language Processing as a foundation of the Semantic Web', *Foundations and Trends in Web Science*, 1, 199-327.

Slavoj Žižek (1997). *The Plague of Fantasies*, London: Verso.