Position Paper for Workshop on Artificial Companions in Society: Perspectives on the Present and Future, Oxford Internet Institute,25th–26th October, 2007
Published in *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, Ed. Yorick Wilks, John Benjamins, pp 179–200.
Revised: April 25, 2010

Requirements for Artificial Companions: It's harder than you think

Aaron Sloman
School of Computer Science,
University of Birmingham, UK
`http://www.cs.bham.ac.uk/~axs/`

Abstract

Producing a system that meets plausible requirements for Artificial Companions (AC.s), without arbitrary restrictions, will involve solving a great many problems that are currently beyond the state of the art in Artificial Intelligence (AI); including problems that would arise in the design of robotic Companions helping an owner by performing practical tasks in the physical environment. In other words, even if the AC is not itself a robot and interacts with the user only via input devices such as camera, microphone, keyboard, mouse, touch-pad, and touch-screen, and output devices such as screen and audio output devices, nevertheless it will, in some circumstances, need the visual competences, the ontology, the representational resources, the reasoning competences, the planning competences, and the problem-solving competences that a helpful domestic robot would need. This is because some of the intended beneficiaries of ACs will need to be given advice about what physical actions to perform, what physical devices to acquire, and how to use such devices. I shall give examples illustrating the need for such competences.

NOTE: some of the text was changed for the version in the book. E.g. 'Digital Companion', and 'DC' were replaced by 'Artificial Companion' and 'AC'. I did not have time to check all the changes before final versions were required. I object strongly to the removal of section numbers to facilitate cross references in a scientific volume. They have been retained in this version. So this version should be regarded as definitive, not the version published in the book.

Contents

1 Introduction

This paper is a response to an invitation to join a discussion on Digital Companions (henceforth DCs – sometimes referred to as "Artificial Companions" or ACs). DCs were described in the invitation as

> "conversational software-based artificial agents that will get to know their owners over a substantial period. These could be developed to advise, comfort and carry out a wide range of functions to support diverse personal and social needs, such as to be artificial companions for the elderly, to help their owners learn or to sustain the fitness and health of their owners."

Producing a system that meets these requirements, without arbitrary restrictions, will involve solving a great many problems that are currently beyond the state of the art in AI, including problems that would arise in the design of robotic companions helping the owner by performing practical tasks in the physical environment. In other words, even if the DC is not itself a robot and interacts with the user only via input devices such as camera, microphone, keyboard, mouse, touch-pad, and touch-screen, and output devices such as screen and audio output devices, nevertheless it will, in some circumstances, need the visual competences, the ontology, the representational resources, the reasoning competences, the planning competences, and the problem-solving competences that a helpful domestic robot would need. This is because some of the intended beneficiaries of DCs will need to be given advice about what physical actions to perform, what physical devices to acquire, and how to use such devices. I shall give examples illustrating the need for such competences.

Moreover, although there are many robot projects aiming to produce useful, more or less intelligent, autonomous robots, and some of them[1] are very impressive at performing limited tasks (e.g. welding car bodies, or moving around on rugged terrain) they are nowhere near meeting the requirements of a helpful multi-purpose companion. That is because there are significant differences in requirements for forms of representation and information-processing architectures between (a) machines that are merely able to perform complex tasks in a physical environment and (b) machines that understand what they have done, what what they could do, what they should not do, why they should not do it, what the consequences of actions will be, what further options could arise if a possible action were performed, how all this relates to what another individual could or should do, and can also communicate some of this to other individuals. Insects are very impressive as regards (a) but appear not to have competences of type (b) except in very limited ways (e.g. communicating through pheromone trails, or bee dances).

One of the problems in producing robots with competences of type (b) is that the requirements for such systems are extremely complex and subtle, and far from obvious, whereas many researchers think that the requirements are obvious and well understood, so that the only task is to work out how to produce systems that meet the requirements. Similar problems will arise for work on DCs.

The assumption that the requirements are clear is based on a failure to understand the rich and varied details of the processes of perceiving, understanding and interacting with all the different sorts of objects and situations that can occur in a normal house, in contexts where different sorts of human needs have to be met, many of which cannot be predicted in advance, so that solutions will require creative problem-solving by the DC at run time. I shall attempt to explain some of those problems briefly below, though more details can be found in (Sloman, 1982, 1989, 1996, 2001b, 2008a, 2009; Sloman, Wyatt, Hawes, Chappell, & Kruijff, 2006; Sloman, 2007b, 2008c). One way of expressing the point is that we do not yet understand what all the problems were that biological evolution solved, in producing designs for infants that can develop into human beings in many different physical and cultural contexts. As John McCarthy (2008) noted, evolution did not solve the problem of producing a robot-like system that starts with no knowledge about the environment but has some completely general, immensely powerful, learning system.

One way to avoid these very demanding requirements is to restrict severely the intended role of the DC, for instance producing only "engaging" DCs (defined below), and only the simpler sorts of "helper" DCs. However, it is important (a) that both researchers in this area and potential users of such devices, and their carers, come to understand the variety of requirements for helpful DCs, (b) that they understand why some of the requirements are very hard to meet, and may remain so for many years, and (c) that people who consider selecting, buying, using, or funding development of DCs are not misled into expecting what cannot be delivered.

Section 2 offers a first draft high level taxonomy of types of *function* that might be desired for DCs, so that we can distinguish functions that might be provided on different time-scales and

[1]E.g. the BigDog and LittleDog robots of Boston Dynamics.

understand which expectations are likely to remain unfulfilled in the next decade or longer.

The following section, 3, offers a shallow taxonomy of types of *motive* that may drive researchers, funding organisations, carers and users, involved in funding, developing, purchasing and using DCs. For ethical reasons we need to distinguish clearly (a) the motives, interests, and needs of the end-users (the people who are to be helped, advised, comforted, entertained, or whatever) and (b) the motives interests and needs of others involved, e.g. carers, relatives, and the companies or organisations responsible for providing care, and also the scientists and engineers involved in developing DCs.

The remaining sections expand on the difficulties in achieving the more ambitious functions. There are ethical issues related to production and use of DCs, but that is not the main topic of this paper, which is more concerned with technical requirements, scientific problems and near-term feasibility. Ethical issues will, however, arise in connection with some of the more sophisticated enabling functions of DCs. The paper ends with a summary and some warnings.

1.1 Terminology: different human roles.

I use the word "user" to refer to individuals for whom DCs are provided. The word "owner" was used in the quotation in Section 1, but we need to distinguish the individuals to be helped by DCs from those who buy and own them, such as helping organisations, relatives of users, etc. A user of a DC will regularly interact directly with the DC, and is expected or intended to derive some benefit from doing so. I shall use the word "carer" to refer to other individuals who are involved in the decision to acquire the DC, or who have, or feel they have, some ongoing responsibility to the user. Typically carers will not interact with the DC, or will do so less frequently than the user. We can distinguish "direct carers" who interact regularly with the user and, where appropriate with the DC, in order to train it, check its logs, arrange repairs, etc., and "indirect carers" who have some responsibility for or concern about the user, but either choose not to be or are unable to be closely involved, so that they leave it to others to be direct carers, and in some cases contribute to the cost of the service. In some cases the roles will overlap: e.g. the same individual might be owner of the DC and direct carer, or owner and indirect carer.

Another group of individuals with interests and responsibilities are the "developers" who work on producing or improving DCs. They may have a variety of different motives (including worrying about what their own quality of life will be in old age!), though most seem to believe they are working to improve the human condition, at least for some humans.

A DC user's interests are not always the main or sole concern of developers, owners, direct or indirect carers or investors funding the development, and this can raise serious ethical issues, for instance when a DC is provided not in order to raise the quality of life of the user, but merely to reduce the burden on a carer, or in order to make money for owners, direct carers and others. In some cases there is no conflict, e.g. where DCs really do enhance the quality of life of users and also provide an income for owners, carers, developers and funders, or reduce burdens on carers. However such ethical issues are not the main topic of this paper.

2 Types of function for digital companions

A great deal of work is currently being done on developing artificial agents (robotic and non-robotic) that can interact with humans, as toys, as game participants, as tutors, as counsellors, as marketing devices, in various forms of entertainment, and for research on various topics, including investigations of affective processes (much of it inspired by Picard (1997)). Not all the devices being developed are robots and not all are being developed as long term digital companions. Many of them will share requirements with DCs. The target functions of DCs and other interactive devices can be roughly divided into two broad categories: (a) 'engaging' functions that focus on the quality of interaction between machine and user and the ability of the machine to hold the attention and interest of the user, keeping the user entertained, amused, or merely interested, independently of any benefit that the user may gain from the interaction; and (b) 'enabling' functions that focus on meeting needs or achieving goals of the user, The division is not very sharp, however. Each of those will be further subdivided, in order to contrast the functions that will be hardest to achieve from the rest.

2.1 Engaging functions

It is useful to identify a collection of functions that many developers currently aim to provide, which I call "engaging functions", e.g. keeping the user happy, quiet, entertained, uncomplaining, etc. This can often be achieved by providing fairly shallow competences in machines, as illustrated by entertaining chatbots, motorised toys, speaking dolls, or in some cases merely cuddly toys. The kinds of devices that can be used to provide engaging functions include at least the following sub-categories, though this is not meant to be a definitive classification:

- TOYS: Entertaining devices for occasional use (compare dolls, simple computer games, recorded music players, and what could be described as "entertaining dummy-humans", i.e. superficially human-like electronic products that have amusing physical behaviours, or provide entertaining verbal or graphical interaction, e.g. some chatbots.)
- INTERACTIVE DEVICES: Products that are intended to be used regularly to provide interest, rich and deep enjoyment through active participation by the user, or a feeling of companionship (compare pets, musical instruments, construction kits, more intellectually demanding computer games).
- DUMMY-HUMANS: Products that are intended to be regarded by a user as being like another caring, feeling individual with whom a long term relationship can develop – even if that is based on an illusion because the machine is capable only of shallow manifestations of humanity: e.g. learnt behaviour rules, such as nodding, smiling, head turning, gazing at faces, making cute noises, or in some cases apparently asking to be comforted, etc.

Unfortunately, there are some theories of emotion that treat emotions as little more than dispositions to produce shallow manifestations that can be recognised as indicating sadness, surprise, being amused, anger, fear, relief etc. Researchers who believe such theories often assume that they are working on giving machines emotions when all they are doing are producing machines that at best either deceive or engage. For a discussion of differences between deep and shallow theories of emotions see (Wright, Sloman, & Beaudoin, 1996; Sloman, 2001a).

The engaging functions have probably attracted the most research and development effort, since they are the easiest to achieve, and, I suspect, most fun for researchers. Some of this work is very superficial, namely getting machines or avatars to produce various mildly convincing imitations of human behaviours, without trying to understand or replicate the information-processing that leads up to and produces, maintains or modulates those behaviours, or reactions to them, in humans or other animals. I worry that some, but not all, work on engaging functions is pursued cynically for financial gain though deception of potential users, owners, carers, or funders, or else pursued naively because the researchers are deceived about the significance of such products, because they believe incorrect theories. (For a first draft attempt at explaining aspects of grief that contradict most theories of emotion, for instance because grief can endure for months or years, and can coexist with short term states such as joy, amusement, relief, etc., see (Wright et al., 1996).) The following sections mostly ignore work on engaging functions, except by contrasting them with enabling functions.

2.2 Enabling functions

Future DCs will be developed in order to produce more sophisticated functions than the engaging functions, for example providing help, assistance, advice, etc. for users, including helping them to do things they would wish to do even if they did not have a DC to help, for instance, moving around, getting dressed, reading newspapers, having a bath or shower, cooking food, learning some new skill, deciding whether to call a doctor, being reminded to take pills, dealing with a minor injury or domestic mishap, obtaining information, etc. Some enabling functions can be part of therapeutic treatment, e.g. helping someone who has been injured or had a stroke do exercises to recover lost abilities, as in (Tapus, Tapus, & Mataric, 2008).

DCs providing enabling functions may have various levels and kinds of competence. The following is a first draft, rather shallow, sub-division into three successively more sophisticated levels. A more substantive analysis would require far more detailed subdivisions, e.g. based on types of competence required for different functions.

- HELPERS:
  These are systems that can reliably provide help and advice that meets practical everyday needs as well as occasional unexpected problems, for example: detecting that the user has fallen and is not moving, then calling an ambulance service; answering questions using a database of frequently asked questions, with answers that are likely to suit all or most questioners; detecting a knock at the door and flashing a light to attract a deaf user's attention; providing advice or answering questions by using an algorithm for searching the internet for such questions and their associated answers, and many more. These systems provide help on the basis of explicitly provided information and algorithms, or as a result of being trained in advance on some corpus to detect classes of problems and associate them with specific answers from the corpus, or in some cases answers derived by simple logical inference or mathematical calculation. Such a system will not develop new capabilities unless re-programmed or re-trained. A helper that is able to extend its capabilities while on the job will fall into one of the following categories.
- DEVELOPING HELPERS:
  These are companions that are capable of developing, over time, an increasingly deep

and broad (perhaps also human-like) understanding of the user's environment, needs, preferences, values, knowledge, and capabilities, so that they extend and improve their ability to help and support the user, as opposed to being restricted by the functionality given to them by their designers. An example might be noticing that the user is becoming more absent-minded and learning to detect situations arising, e.g. when tasks are left unfinished, when the user needs a reminder or warning, etc., and learning how to provide such advice and help in a manner that is not resented or disliked. Another example might be noticing that the user is starting to find certain physical tasks increasingly difficult, detecting a pattern in the tasks and the problems, and reporting the discovery to a carer.

How such learning can occur, what kind of information-processing architecture can support it, what physical interfaces are required (e.g. sensors distributed around the home) and how much the user has to cooperate if the learning process is to be reliable, are all important details that will not be discussed here.

Learning helpers are needed because it will be impossible for designers to anticipate everything and also because the user's situation and the user's needs and preferences will change over time. More advanced abilities to extend themselves will be required in DCs in the next category.

- CARING DEVELOPING HELPERS:
  A human user and the user's environment will provide very many opportunities for learning, since over time the user may change (e.g. deteriorating physical and/or mental health, loss of a particularly close human carer, changed financial circumstances, buildings, furniture, or appliances deteriorating, or being replaced by new versions with different characteristics, acquisition of new household appliances, and development of new interests, hobbies and concerns by the user).

  The DC will have to choose which of these changes in its environment to attend to, and make decisions about which observations are relevant to its function. How to control such a process is far from obvious, and it is likely to be very difficult. One form of control would be to make the helper really want to do what is best for the user. This is analogous to biological parents wanting to do what is best for their offspring instead of merely having fixed reactions to situations involving their offspring, selected either by evolution or some social pressure group.

  So the most sophisticated DCs may need to be developing helpers that grow to care about the user and really want to help when things go wrong or have a significant probability of going wrong, and which want to find out how best to provide such help.[2]

  This sort of caring (real, not simulated) will be required to ensure that the DC is motivated to find out what the user wants and needs, and so that it can deal with conflicting criteria in ways that serve the user's interests.

  Such a helper will want to find out what the user wants, prefers, likes and dislikes; and will also come to want to act in accordance with those attitudes. It may sometimes notice

[2]In Sloman (1978) an argument was presented that it is possible for machines to have their own goals, preferences and values if, like humans, instead of being given only *fixed* high level goals, they start off with mechanisms for absorbing, generating, and modifying their goals, preferences and values during development. When such machines have their own goals, treating them as slaves, as proposed by one of the workshop participants, would be highly unethical, like racism.

conflicts between what the user wants and what is good for the user and will generally favour the long term benefit of the user,[3] making exceptions in the sorts of cases where a caring human helper would.

There is a pattern in these three broad categories that is also relevant to the successively more demanding problems “solved” by evolution in producing increasingly cognitively and architecturally sophisticated organisms to accommodate increasingly complex, varied and unpredictable environmental demands and challenges (as discussed in Sloman & Chappell, 2005; Chappell & Sloman, 2007; Sloman, 2007a).

These divisions are not unique to DCs, and the divisions are not very well defined: they merely provide a crude, first draft, classification, relevant to the purposes of this paper. Each category can be broken down into further sub-categories, and each group of categories probably needs to be extended by additional members not included in the existing divisions.

I have separated the engaging functions from the enabling functions in order to indicate that the main focus of this paper is the enabling functions and how to provide them. I am interested only in the problems of producing helpers, especially caring, developing helpers, and not at all interested in producing engagers. Unfortunately, producing good helpers, especially caring, developing helpers, is much harder than producing shallow engagers, and, equally unfortunately, it seems that some researchers ignore the differences, and consequently underestimate the difficulties in producing the enabling DCs.

Later sections will examine some enabling functions of DCs more closely, explaining why, for those functions, major advances are needed in both our understanding of requirements and in our knowledge of how to design and implement machines with human-like competences, as opposed to merely giving the appearance of having such competences through superficial behaviour in constrained testing conditions. Achieving the ambitious enabling functions will require us to solve interdisciplinary research problems involving AI, philosophy, psychology, linguistics, biology and possibly neuroscience. That includes acquiring a deeper understanding of the problems solved by evolution in producing humans and other animals. Recent work by developmental psychologists is relevant, though unfortunately not all of them adopt what McCarthy calls “the designer stance” (McCarthy, 2008).

3 Motives for developing, funding, buying or using DCs

In addition to distinguishing the different kinds of *function* that can be served by a DC we can ask about the *reasons* or *motives* behind decisions to acquire one, bearing in mind that sometimes the user will not be the decision-maker. Decisions to apply such systems, and to choose between different combinations of functions to go into a DC will often not be made by the user, or not by the user alone. For example, government or medical agencies, members of the family, and managers of homes for such people may all be involved, either acting alone or in cooperation with one another. They may or may not take the user’s needs, desires and preferences into account, and when they do take them into account they can give the user’s needs etc., differing relative weights.

In particular there are motives that concern only the user’s desires and benefits, and motives of other humans who care about or have responsibilities to the user. Example motives for providing a DC are:

[3] Some of the issues involved in this were discussed in Will Lowe’s contribution to the workshop.

- Because the user wants the DC:
  For example, the user may sincerely prefer to be helped by a DC so as not to have to impose on other humans and the others involved may respect that preference, even if they would prefer to provide the care themselves.
- Because carers have constraints:
  The others may want the DC to be available to fill gaps and provide needed help and care when human carers are unavoidably unavailable, e.g. because they have children they have to look after, or because they need to go to work to earn funds to pay for the care, or because they are badly needed elsewhere, etc.
- Because carers don't care enough:
  The others may wish to use the DC in order to enable them to avoid tasks that they find distasteful or because they have other personal preferences/priorities
- Because carers have theories about needs of users:
  A particular type of DC, like a type of medicine, or a type of home for a user, may be chosen by carers because they have theories about what would be "best" for the user. Such theories may or may not be accepted by the users in whose alleged interest such decisions are taken. Moreover, even if a theory is accepted by everyone involved, it may be fashionable but false.

Obviously these goals and preferences can in some cases be somewhat cynical or selfish: The main beneficiaries of a DC in some situations will not be the user but others connected with the user, either because of personal relationships or because of contractual relationships (e.g. the owners of nursing homes, or retirement homes). The different motives are not necessarily all sharply distinguishable. There may be fuzzy intermediate cases, including mixed motives. I have mentioned motives because they are important when considering the ethics of selection and use of DCs, but will not discuss them further.

4 Problems of achieving the enabling functions

I work on robotics, not in order to produce useful machines, but because that is a very good way of addressing many old philosophical problems (Sloman, 1978), and also because it is a good way to attempt to understand some of the problems solved by biological evolution. It is also possible for such research to produce practically useful designs. However, I have no interest in making or using machines with the "engaging" functions described previously, i.e. those described earlier as: toys, engagers and dummy-humans. I have no objection to others building these things – for the right motives, though I don't think I would ever want to use them myself. E.g. I intensely dislike pseudo-human interfaces, with smiling, nodding, expressions, moving eyes, "emotional" voices, and a stock of linguistic stratagems for expressing concern, interest, or understanding, whether what is expressed exists or not. I hope nobody ever expects me to put up with such things if I become disabled enough to need a DC. The widespread annoyance produced by Microsoft's "helpful" paperclip suggests that I am not alone. Unfortunately, researchers, especially if they are engineers rather than experienced psychologists, often assume for bad reasons that it is obvious that certain kinds of interface will be preferred by users.

The rest of this paper is concerned only with a subset of the problems of producing DCs with enabling functions. Really useful general-purpose DCs are very difficult and way out of reach in the foreseeable future, for reasons that do not seem to be widely understood and which I shall try to explain.

The arguments are not the same as those used by opponents of AI e.g. (Dreyfus, 1979; Searle, 1980; Weizenbaum, 1976): I am *not* claiming that it is impossible to produce human-like machines, or that human-like machines cannot be based on digital computers. I am claiming that the products of evolution and individual human development include many competences that have a depth and complexity that needs to be understood much better if we are to produce really good digital companions of the enabling type. It will be much harder to achieve the enabling functions than most people realise, because it will be necessary to provide some of those competences that are not yet understood. The *currently* most popular AI techniques for competent machines, making heavy use of data-mining and statistical learning, are inadequate for the task.

The main point is that the detailed requirements for DCs to meet the enabling/helping specifications are not at all obvious, and have implications that make the design task very difficult in ways that have not generally been noticed. Building such systems will require a deep new understanding of some hitherto unexplained human competences. Perhaps they will eventually be achieved if we analyse the problems properly.

4.1 Kitchen mishaps

Many of the things that crop up will concern physical objects and physical problems. Someone I know knocked over a nearly full coffee filter close to the base of a cordless kettle. This caused the residual current device in the fuse box under the stairs to trip, removing power from many devices in the house. Fortunately she knew what to do, unplugged the kettle and quickly restored the power. However, she was not sure whether it was safe to use the kettle after draining the base, and when she tried it later the RCD tripped again, leaving her wondering whether it would ever be safe to try again, or whether she should buy a new kettle. In fact it proved possible to open the base, dry it thoroughly, then use it as before.

Should a DC be able to give helpful advice in such a situation? Would linguistic interaction suffice? How? Will cameras and visual capabilities be provided? People who work on language understanding often wrongly assume that providing 3-D visual capabilities will be easier, whereas very little progress has been made in understanding and simulating human-like 3-D vision and spatial understanding, which involves far, far more than recognising things. Human vision includes a wide and deep collection of competences, and ontologies, some analysed in (Sloman, 2008a). Many researchers confuse seeing with recognising, which is wrong because you can see something, and even interact with it, without recognising it, e.g. by pushing it out of your way or smashing it with a sledgehammer. Moreover, even when something is recognised either as a type of entity or as a previously encountered individual, there is more to seeing than that. For instance, as Gibson pointed out, seeing and other forms of perception provide information about positive and negative affordances, and such perception usually builds on the more basic competence of perceiving which processes are occurring or can occur in a situation. As argued in (Sloman, 2008a), that requires the use of forms of representation that can encode information about processes occurring in the environment, including processes that involve multiple concurrent changes, e.g. of shape, position, orientation, and relationships between objects, some metrical, some semi-metrical[4] some topological, some functional, and some causal. It is also useful in many

[4]E.g. A is closer to B than C is to D.

contexts to understand the behaviours of different kinds of material, e.g. fur, skin, porridge, metal, plastic, wood, water, detergent, powders of various sorts, paper, cloth, cardboard, cotton wool, etc., and at present no machines come close. As far as I know the problem of giving machines an understanding of how shape and materials interact in the production of behaviours has barely been recognized, let alone solved.

There are no machines that have such capabilities, partly because only a very narrowly restricted set of requirements for vision has been studied by researchers in vision and robotics, and partly because it is very hard to produce appropriate designs, for reasons developed in more detail in (Sloman, 2008a). Moreover some of the forms of representation and processing capabilities are needed for thinking about spatial structures and processes even when they are not currently perceived, for instance, when planning actions, or working out what an opponent might do but is not yet doing.

For these reasons, it is likely that many of the functions of a DC that can "advise, comfort and carry out a wide range of functions to support diverse personal and social needs, such as to be artificial companions for the elderly, to help their owners learn or to sustain the fitness and health of their owners" will require them to have access to cameras and other sensors and to be able to reason about practical problems involving spatial structures and processes. It is not always noticed that these capabilities are required even for purely linguistic interactions. E.g. the person who spilt coffee over a kettle base told me what had happened over the phone and we were able to discuss possible strategies because I could think and reason about the physical configurations involved without actually seeing them. When I later, out of curiosity, attempted to use google to find advice about what to do in such a situation, the closest things I found were advertisements for kettles with sealed bases that were protected against spills.

It is not always appreciated how much humans learn about space, time, structures, processes and causal interactions in the first few years of life, providing a deep and general understanding of many features of the environment on which a wide range of later developments can build. If DCs, whether robots or not, are to emulate the depth of understanding, creativity and flexibility that can be based on those spatial competences we shall need far more progress in designs suited for use in robots which can also be used in disembodied DCs.

It is also not widely appreciated that recent trends in AI and cognitive science that disparage symbolic AI and emphasise the role of embodiment and the dynamical interactions between bodies and the environment, e.g. (Brooks, 1990, 1991) leave out of account many human cognitive functions that involve thinking about, reasoning about and planning actions in environment, which require forms of representation and architectures that are different from those needed for online interaction, as explained in (Sloman, 2009, 2008c).

4.2 Identifying affordances and searching for things that provide them

Understanding a human need and seeing what is and is not relevant to meeting that need may require creative recombination of prior knowledge and competences.

Suppose an elderly user finds it difficult to keep his balance in the shower when soaping his feet. He prefers taking showers to taking baths, partly because showers are cheaper. How should the DC react to this problem? Should it argue for the benefits of baths? Should it send out a query to its central knowledge base asking how people should keep their balance when washing their feet? (It might get a pointer to a school for trapeze artists, or a balancing pole suitable for tight-rope walkers.)

What if the DC designer had not anticipated that problem, and nothing suitable was in the latest knowledge updates for the DC? What are the requirements for the DC to be able to invent the idea of a folding seat attached to the wall of the shower, that can be temporarily lowered to enable feet to be washed safely in a sitting position? Alternatively what are the requirements for it to be able to pose a suitable query to a search engine? How will it know that safety harnesses and handrails are not good solutions? Of course, once it had understood that a folding shower seat might be a good solution, it might use that to direct searches for local suppliers of shower seats. If it found none it might have to dream up an alternative solution so as to be able to decide what to search for on the internet. There are many more examples of issues that can arise in a domestic situation where individuals of various sorts need help and advice, requiring the helper to think about interacting 3-D structures and processes.

Another example: I expect very few people reading this have ever spilt a bowl of porridge on a carpet. Yet you probably know quite a lot about carpets in general (or a particular type of carpet, e.g. cord, mohair, wilton, etc.), about physical qualities of porridge, and about cleaning messes, and you can combine what you know to give to give specific advice about a novel situation. For example that will rule out certain approaches, such as using a vacuum cleaner, a broom, a bucket of water and a mop. Ruth Aylett responded to the example by pointing out that someone carrying a bowl of porridge may also be carrying a spoon, which would be one of the best devices to use, in conjunction with the bowl, to start cleaning up the mess.

A DC could be pre-programmed with many specific pieces of advice to offer in specific situations, but providing the ability to deal creatively with novel problems by combining existing competences and knowledge in a novel way, is much harder, and requires the DC to have a much deeper understanding of space, time, causation, and interactions between different types of matter and the various shapes in which they can occur and how the combination of matter and shape can influence the processes that are possible.

4.3 Remembering particularities: episodic memory

Another important human competence that no AI system already has includes the construction of so-called "episodic memories", spanning different regions of space and time at different scales. This involves learning not only new concepts and new generalisations, but also acquiring and storing information about particular places and times, and at least sometimes being able to recall such particularities when they are relevant to a new problem. For example, as you move around a room, or various parts of a building, or a town, you can at any time see only a small spatial region, yet you remain aware that that there are things you have previously seen or read about or been told about that exist but are out of sight, and you are also aware of things you have done and other happenings that occurred recently, and which no longer have any impact on the senses. Some robots that can do SLAM, i.e. Simultaneous Localisation and Mapping, have a special case of that competence, though usually the information stored, and what can be done with it is very limited.

This information about what exists un-sensed now or existed in the past, or may exist in the future, has many applications, including planning routes and actions, describing the environment or recent events to others, and also understanding what other speakers are referring to when they talk about the things or past happenings in the environment. So these spatial competences play an essential role in human linguistic competences, of kinds that the DC is likely to need.

Giving machines an understanding of physical and geometrical shapes, processes and causal interactions of kinds that can occur in an ordinary house is currently far beyond the state of the art. Compare the 'Robocup@Home' challenge, still in its infancy (`http://www.ai.rug.nl/robocupathome/`). Major breakthroughs of unforeseen kinds will be required for progress to be made, especially breakthroughs in vision and understanding of 3-D spatial structures *and processes*. Some simple examples of the required competences are discussed in an online discussion paper on "Predicting Affordance Changes".[5] Of course, one response to all this would be to aim for much simpler and more easily attainable competences, such as simply providing the engaging functions. Either that, or implementing fairly simple helper packages is probably what many providers of DCs will aim for in the near future. I shall ignore such goals in the rest of this paper and focus on requirements for addressing the harder problems.

4.4 More abstract problems

So far all the problems presented relate to a richly varied 3-D environment in which structures and processes occur that need to be understood, enabled, prevented, predicted, explained, coped with, undone, and so on, which have so far defeated AI systems except for very special, simple cases, although many of the required competences are found in young children and non-human animals.

There are also meta-semantic competences that young humans develop, namely the ability to represent some things, including themselves, as information-processors, and to develop and use explicit or implicit theories about states and processes that can occur in information processors, such as having and using concepts and being able to understand some things while failing to understand others. Some meta-semantic competences involve 3-D structures and processes, for instance the ability to infer the intention in an observed action, but not all do, for instance, wondering whether someone understands long division. Many, though not all, uses of meta-semantic competences involve using language to describe mental states and processes, of oneself or others. Although the problem of getting a machine to handle referential opacity has been recognised for a long time (e.g. see McCarthy & Hayes, 1969) there is no agreement on how to solve the problem. Some researchers favour using special kinds of modal logic, or other notational devices. I suspect that the best solution is to use an information-processing architecture that supports encapsulation, e.g. an architecture that allows beliefs and inferences of other individuals to be simulated, which need not involve treating accepting them as one's own.

Coping with referential opacity is one among many problems that arise when using language to communicate. Another competence that interfaces with the ability to acquire and store information about the physical environment, including portions that are not in view, and events that occurred in the past, is the ability to understand referring expressions, including indexicals, proper names and definite descriptions (e.g. "The mug I used yesterday"). Machines that cannot build human-like information structures referring to the spatial and temporal environment including currently unobservable portions will not be able to resolve ambiguities and interpret all referring expressions.

[5]Available online at `http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0702`

4.5 Is training the solution?

Many AI researchers have addressed problems of communication by machines, using natural language (written, spoken, or encoded in other media), gestures, facial expressions, bodily postures, etc. However, much of this work has involved either providing shallow behavioural rules directly, or else enabling machines to be trained to produce different responses to different situations by being presented with large numbers of examples.

This may work for simple forms of communication in a restricted range of contexts, but for more general human-like understanding of behaviours and communications that are not all instances of a fixed class of stereotypes, the understander needs a richer, deeper theory about the kind of thing a communicator can be, including a theory about the information processing architecture that lies behind expressions of beliefs, intentions, desires, curiosity, amusement, interest, boredom, disgust, pleasure, regret, shame, embarrassment, annoyance, indignation, surprise, puzzlement, wanting to understand, misunderstanding, incomprehension, refusal, acceptance, willingness, resigned acquiescence, suspicion, attending, noticing, ignoring, expecting, remembering, forgetting, preference, mood, weariness, jauntiness, alertness, sleepiness, being intoxicated, and many more.

4.6 Beyond behavioural dispositions

These states and processes may or may not manifest themselves directly in behaviour, but they are not pure behavioural dispositions, since they depend on internal structures and processes in a rich and complex information processing architecture, including an architecture that supports meta-semantic competences used both in self-understanding and in other-understanding. The behaviours are the outcome of interactions between components in the architecture. We still do not understand precisely what variety of architectures humans can have, and what sorts of implicit theories individuals develop about their own and others' architectures that enable them to get on in life with only partial understanding of the enormous complexity of other people, though there are many disciplines investigating this from different viewpoints, using different ontologies and methodologies, including conceptual analysis in philosophy (Ryle, 1949).

Much research in AI ignores relevant work in other disciplines, including philosophy, and because much of the work is done by people trained in electronics, engineering, computer science, mathematics, etc. and not in linguistics, psychology, or philosophy, this leads to over-simplified models. There are attempts to explore more of the complexity required, e.g. (Minsky, 2006), but the work is still at a very early stage. We therefore cannot expect DCs in the next few decades to be able to do more than produce shallow responses on the basis of shallow understanding of what they perceive.

This may work well enough in situations where the human-machine interaction is constrained to a very narrow task, such as booking train tickets, asking directions in a department store, filling in an order form, enquiring about legal rights or a certain medical condition, for example. But the situations envisaged for DCs as described in the invitation to contribute this paper will not be so constrained, since DCs will have to go into different homes, with different sorts of people whose lives, loves, memories, capacities, incapacities and requirements can vary enormously, and also change over time.

Examples given previously illustrate the need for creative understanding of physical problems.

There will also be a need for creative understanding of thought processes, cognitive competences, preferences, motivations and values. There is space here only for a few examples. Suppose the user U has decided to compile a database of information about his relatives and close friends and asks the DC to help with the task. Suppose U is an intense atheist, and while trawling for information about U's siblings the DC finds that U's brother has written a blog entry supporting creative design theory, or discovers that one of U's old friends has been converted to Islam and is training to be a Mullah. How should the DC react? Compare discovering that the sibling has written a blog entry recommending a new detective novel he has read, or discovering that the old friend is taking classes in cookery. What about getting evidence suggesting that U's spouse had been having an affair with a friend long ago? Could the DC reason that telling U about it might assuage guilt about an affair U had once had?

A DC may need the ability to take account of emotional responses that news items may produce. How will it work out when to be careful when reporting new information? Where will the DCs goals, preferences, and values come from? (More on that later.)

5 Is the solution statistical?

The current dominant approach to developing language understanders and advice givers involves mining large corpora using sophisticated statistical pattern extraction and matching. This is much easier than trying to develop a structure-based understander and reasoner, and can give superficially successful results, depending on the size and variety of the corpus and the variety of tests. But the method is inherently broken because as sentences get longer, or semantic structures get more complex, or physical situations get more complex, the probability of encountering recorded examples close to them falls very quickly. Then a helper must use deep general knowledge to solve a novel problem creatively, often using non-linguistic context to interpret many of the linguistic constructs (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0605). Some machines can already do creative reasoning in restricted domains, e.g. planning and mathematical reasoning, but they are still very limited.

5.1 Why do statistics-based approaches work at all?

The behaviour of any intelligent system, or collection of individuals, will leave traces that may have re-usable features, and the larger the set the more re-usable items it is likely to contain – up to a point. For instance it may not provide items relevant to new technological, or cultural developments or to highly improbable but perfectly possible physical configurations and processes. So any such collection of traces will have limited uses, and going beyond those uses will require something like the power of the system that generated the original behaviours.

In humans (and some other animals), there are skills that make use of deep generative competences whose application requires relatively slow, creative, problem solving, e.g. planning routes. But practice in using such a competence can train powerful associative learning mechanisms that compile and store many partial solutions matched to specific contexts (environment and goals). As that store of partial solutions (traces of past structure-creation) grows, it covers more everyday applications of the competence, and allows fast and fluent responses.

Some statistical AI systems that cannot generate the data can infer those partial solutions from large amounts of data. But because the result is just a collection of partial solutions it will always have severely bounded applicability compared with humans, and will not be extendable in the way human competences are, namely by creatively recombining things known previously so that solutions to problems can be derived from the combination, as in designing a new type of machine made of familiar parts.

Moreover, if trained only on text a machine will have no comprehension of non-linguistic context. Dealing with novel problems and situations requires different mechanisms that support creative development of novel solutions.

If the deeper, more general (and slower) competence is not available, wrong extrapolations can be made, inappropriate matches will not be recognised, new situations cannot be dealt with properly and further learning will be very limited, or at least very slow. In humans the two systems work together to provide a combination of fluency and generality. (Not just in linguistic competence, but in many other domains.)

Occasionally I meet students who manage to impress some of their tutors because they have learnt masses of shallow, brittle, superficially correct patterns that they can string together – without understanding what they are saying. They function like corpus-based AI systems: Not much good as (academic) companions.

6 Can it be done?

Producing a DC of the desired type may not be impossible, but is much harder than most people realise and cannot be achieved by currently available learning mechanisms. (Unless there is something available that I don't know about). Solving the problems will include:
(a) Learning more about the forms of representation and the knowledge, competences and meta-competences present in prelinguistic children who can interact in rich and productive ways with many aspects of their physical and social environment, thereby continually learning more about the environment, including substantively extending their ontologies. Since some of the competences are shared with other animals they cannot *depend* on human language, though human language depends on them. However we know very little about those mechanisms and are still far from being able to implement them.

(b) When we know what component competences and forms of representation are required, and what sorts of biological and artificial mechanisms can support them, we shall also have to devise a *self-extending architecture* which combines them all and allows them to interact with each other, and with the environment in many different ways, including ways that produce growth and development of the whole system, and also including sources of motivation that are appropriate for a system that can take initiatives in social interactions. No suggestions I have seen for architectures for intelligent agents, come close to requirements for this, though there are many AI projects taking small steps that could be useful.

I suggest the only reliable way to meet the objectives is to understand and replicate, and later on to build on, some of the generic capabilities of a typical young human child, including the ability to want to help.

6.1 What's needed

Before human toddlers learn to talk they have already acquired deep, reusable structural information about their environment and about how people work. They cannot talk but they can see, plan, be puzzled, want things, and act purposefully. In the first two or three years of life they are constantly learning about large numbers of affordances related to various kinds of processes involving objects and situations in the environment. They are also learning about epistemic affordances – learning to discern situations that do and do not provide good task-specific information.

They have something to communicate about. That pre-linguistic competence grows faster with the aid of language, but must be based on a prior, internal, formal 'linguistic' competence using forms of representation with structural variability and (context-sensitive) compositional semantics. Jackie Chappell and I have begun to call these forms of representation, which can occur either inside the head or in some public communication "Generalised languages" or "GL"s in (Sloman & Chappell, 2007; Sloman, 2008b).

The use of GLs enables very young children to learn any human language and to develop in many cultures. DCs without a similar pre-communicative basis for their communicative competences are likely to remain shallow, brittle and dependent on pre-learnt patterns or rules for every task.

Perhaps, like humans (and some other altricial species), they can escape these limitations if they start with a partly 'genetically' determined collection of meta-competences that continually drive the acquisition of new competences building on previous knowledge and previous competences: a process that continues throughout life. The biologically general mechanisms that enable humans to grow up in a very wide variety of environments, are part of what enable us to learn about, think about, and deal with novel situations throughout life. Very little is understood about these processes, whether by neuroscientists, developmental psychologists or AI researchers, and major new advances are needed in our understanding of information-processing mechanisms. Some pointers towards future solutions are in the online presentations listed below.

6.2 Alternatives to canned responses

The kitchen mishap, shower support problem, and spilt porridge problem were just three examples among a vast array of possibilities that a domestic DC may need to help with. Of course, if the designer anticipates such accidents and problems, the DC will be able to ask a few questions and spew out relevant canned advice, and even diagrams showing how to open and dry out the flooded base.

But suppose designers had not had that foresight: What would enable the DC to give sensible advice? If the DC knew about electricity and was able to visualise the consequences of liquid pouring over the kettle base, it might be able to use a mixture of geometric and logical reasoning creatively to reach the right conclusions. It would need to know about and be able to reason about spatial structures and the behaviour of liquids. Although Pat Hayes described the 'Naive physics' project decades ago, it has proved extremely difficult to give machines the kind of intuitive understanding required for creative problem-solving in novel physical situations. In part that is because we do not yet understand the forms of representation humans (and other animals) use for that sort of reasoning.

A DC lacking similar mechanisms and a similar deep understanding of our environment may cope over a range of circumstances that it has been trained or programmed to cope with – and then fail catastrophically in some novel situation. Can we take the risk? Would you trust your child with one?

7 Conclusion

A lot of research is being done in connection with the long term goal of developing digital companions, some of which may provide important enhancements for the quality of life especially in a population with ever increasing life-expectancy that does not always go with full physical and mental health in later years. However, there are risks in this research. One such risk is exaggerating the likely value of some of the easier research based on producing shallow behavioural competences in robots and non-robotic systems. The fact that the design goals of such projects are met carries no guarantee that the design goals are worth pursuing. Many of them will be of some use in providing entertainment and fun, the "engaging functions" of DCs, without necessarily providing adequate "enabling functions".

Another risk is underestimating the difficulties of producing digital companions that are able to provide the enabling functions, i.e. able to be helpers or even caring and developing helpers. Progress in providing those functions will require progress in meeting more general goals of AI research, including some goals that are required for producing useful general purpose domestic robots. Section 5.1 argued that despite the early promise of designs based on use of statistical methods applied to large training sets, such designs are inherently limited by the "training envelope" and will not be capable of creative problem solving in dealing with novel situations.

That second difficulty can be tackled by doing far more analysis of the nature of the environment than is usual in AI research, which tends to focus more on mechanisms and architectures, on the assumption that good designs will work in any environment. It seems likely however, for reasons I have tried to illustrate without giving any proofs, that progress in this area will require us to understand what sort of ontology a DC needs to start with, how that ontology can develop over time, including changing in response to changes in the environment, along with developing good forms of representation and architectures to allow information using such ontologies to be acquired, stored, modified, generalised, de-bugged, and used creatively in solving new problems. In particular, this will require machines to be given much better ways of representing processes that can occur in the environment including processes involving interactions between parts of objects, and different kinds of material. Contrary to current popular views this does not require the DC to share a human body morphology, though it does require information processing of kinds that occur in human minds, only a small subset of which is concerned with direct interaction with the environment. The topic is large and complex, and I have given pointers to online presentations and papers exploring the issues in more detail.

7.1 Ethical issues

There are several ethical issues related to the design, selection, and use of DCs. Section 3 distinguished different roles in which people involved with DCs could have different motives, including designers, sellers, purchasers, carers of various sorts, and end-users.

Although many researchers assume that DCs will simply be machines that have no more ethical status than any other domestic appliance, it is arguable for reasons given in Section 2.2 that a DC caring for one or more individuals over a long time period in which many things change, will need to be able to have its own desires, including wanting to help the users, if it is to be successful caring companion, including meeting needs that were not anticipated when the DC was originally designed, and including dealing with cases where there are conflicts between what the user claims to want and what would be in the user's best interests.

If providing effective companionship requires intelligent machines to be able to develop their own goals, values, preferences, attachments etc., including really *wanting* to help and please their owners, then if some of them develop in ways we don't intend, will they not have the right to have their desires considered, in the same way as children do if they develop in ways their parents don't intend? This was discussed briefly in the epilogue to (Sloman, 1978).[6] It can also be argued that Asimov's laws of robotics are immoral, because they are unfair to future robots which may have their own preferences, desires and values.[7] The cavalier suggestion that interests of future machines, no matter how sophisticated they become, can always be disregarded, or at least given lower priority than the interests of human beings seems to indicate an assumption that humans are necessarily uniquely important among information processing systems, an assumption of human ethical centrality that should, in the long run, go the way of the assumption that humans inhabit the centre of the physical universe.

References

Brooks, R. A. (1990). Elephants Don't Play Chess. *Robotics and Autonomous Systems*, *6*, 3-15. (http://people.csail.mit.edu/brooks/papers/elephants.pdf)

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Chappell, J., & Sloman, A. (2007). Natural and artificial meta-configured altricial information-processing systems. *International Journal of Unconventional Computing*, *3*(3), 211–239. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0609)

Dreyfus, H. (1979). *What computers can't do*. New York: Harper and Row. ((revised edition))

Jablonka, E., & Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge MA: MIT Press.

McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, *172*(18), 2003-2014. (http://www-formal.stanford.edu/jmc/child.html)

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of AI. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 4* (pp. 463–502). Edinburgh, Scotland: Edinburgh University Press. (http://www-formal.stanford.edu/jmc/mcchay69/mcchay69.html)

Minsky, M. L. (2006). *The Emotion Machine*. New York: Pantheon.

Picard, R. (1997). *Affective computing*. Cambridge, MA; London, England: MIT Press.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

[6] Available here: `http://www.cs.bham.ac.uk/research/projects/cogaff/crp/epilogue.html`

[7] See `http://www.cs.bham.ac.uk/research/projects/cogaff/misc/asimov-three-laws.html`

Searle, J. (1980). Minds brains and programs. *The Behavioral and Brain Sciences*, *3*(3). ((With commentaries and reply by Searle))

Sloman, A. (1978). *The computer revolution in philosophy*. Hassocks, Sussex: Harvester Press (and Humanities Press). (http://www.cs.bham.ac.uk/research/cogaff/crp)

Sloman, A. (1982). Image interpretation: The way ahead? In O. Braddick & A. Sleigh. (Eds.), *Physical and Biological Processing of Images (Proceedings of an international symposium organised by The Rank Prize Funds, London, 1982.)* (pp. 380–401). Berlin: Springer-Verlag. (http://www.cs.bham.ac.uk/research/projects/cogaff/06.html#0604)

Sloman, A. (1989). On designing a visual system (towards a gibsonian computational model of vision). *Journal of Experimental and Theoretical AI*, *1*(4), 289–337. (http://www.cs.bham.ac.uk/research/projects/cogaff/81-95.html#7)

Sloman, A. (1996). Actual possibilities. In L. Aiello & S. Shapiro (Eds.), *Principles of knowledge representation and reasoning: Proceedings of the fifth international conference (kr '96)* (pp. 627–638). Boston, MA: Morgan Kaufmann Publishers. (http://www.cs.bham.ac.uk/research/cogaff/96-99.html#15)

Sloman, A. (2001a). Beyond shallow models of emotion. *Cognitive Processing: International Quarterly of Cognitive Science*, *2*(1), 177-198.

Sloman, A. (2001b). Evolvable biologically plausible visual architectures. In T. Cootes & C. Taylor (Eds.), *Proceedings of British Machine Vision Conference* (pp. 313–322). Manchester: BMVA.

Sloman, A. (2007a). Diversity of Developmental Trajectories in Natural and Artificial Intelligence. In C. T. Morrison & T. T. Oates (Eds.), *Computational Approaches to Representation Change during Learning and Development. AAAI Fall Symposium 2007, Technical Report FS-07-03* (pp. 70–79). Menlo Park, CA: AAAI Press. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0704)

Sloman, A. (2007b). *A First Draft Analysis of some Meta-Requirements for Cognitive Systems in Robots* (No. COSY-DP-0701). (Contribution to euCognition wiki, also available as, http://www.cs.bham.ac.uk/research/projects/cosy/papers/#dp0701)

Sloman, A. (2008a). Architectural and representational requirements for seeing processes, proto-affordances and affordances. In A. G. Cohn, D. C. Hogg, R. Möller, & B. Neumann (Eds.), *Logic and probability for scene interpretation.* Dagstuhl, Germany: Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

Sloman, A. (2008b). *Evolution of minds and languages. What evolved first and develops first in children: Languages for communicating, or languages for thinking (Generalised Languages: GLs)?* (Research Note No. COSY-PR-0702). Birmingham, UK. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#pr0702)

Sloman, A. (2008c). The Well-Designed Young Mathematician. *Artificial Intelligence*, *172*(18), 2015–2034. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0807)

Sloman, A. (2009). Some Requirements for Human-like Robots: Why the recent over-emphasis on embodiment has held up progress. In B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, & K. Doya (Eds.), *Creating Brain-like Intelligence* (pp. 248–277). Berlin: Springer-Verlag. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0804)

Sloman, A., & Chappell, J. (2005). The Altricial-Precocial Spectrum for Robots. In *Proceedings IJCAI'05* (pp. 1187–1192). Edinburgh: IJCAI. (http://www.cs.bham.ac.uk/research/cogaff/05.html#200502)

Sloman, A., & Chappell, J. (2007). Computational Cognitive Epigenetics (Commentary on (Jablonka & Lamb, 2005)). *Behavioral and Brain Sciences*, *30*(4), 375–6. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0703)

Sloman, A., Wyatt, J., Hawes, N., Chappell, J., & Kruijff, G.-J. M. (2006). Long Term Requirements for Cognitive Robotics. In *Cognitive Robotics: Papers from the 2006 AAAI Workshop: Technical Report WS-06-03, http://www.aaai.org/Library/Workshops/ws06-03.php* (pp. 143–150). Menlo Park, CA: AAAI Press. (http://www.cs.bham.ac.uk/research/projects/cosy/papers/#tr0604)

Tapus, A., Tapus, C., & Mataric, M. J. (2008). User-Robot Personality Matching and Robot Behavior Adaptation for Post-Stroke Rehabilitation Therapy. *Intelligent Service Robotics Journal, Special Issue on Multidisciplinary Collaboration for Socially Assistive Robotics*, *1*(2), 169–183.

Weizenbaum, J. (1976). *Computer power and human reason: From judgement to calculation*. San Francisco: W.H.Freeman.

Wright, I., Sloman, A., & Beaudoin, L. (1996). Towards a design-based analysis of emotional episodes. *Philosophy Psychiatry and Psychology*, *3*(2), 101–126. (http://www.cs.bham.ac.uk/research/projects/cogaff/96-99.html#2)