Relational agents: A critical review.

Robert H. Campbell  
*University of Bolton, R.H.Campbell@bolton.ac.uk*

Mark Grimshaw  
*University of Bolton, m.n.grimshaw@bolton.ac.uk*

Gill Green  
*University of Bolton, G.Green@bolton.ac.uk*

Digital Commons Citation  
http://digitalcommons.bolton.ac.uk/gcct_journalspr/1

This Article is brought to you for free and open access by the School of Games Computing and Creative Technologies at UBIR: University of Bolton Institutional Repository. It has been accepted for inclusion in Games Computing and Creative Technologies: Journal Articles (Peer-Reviewed) by an authorized administrator of UBIR: University of Bolton Institutional Repository. For more information, please contact ubir@bolton.ac.uk.
Relational Agents: A Critical Review

Robert H. Campbell*, Mark N. Grimshaw and Gillian M. Green

Senior Lecturer in Computing and Networks, University of Bolton. Deane Road, Bolton BL3 5AB UK; Reader in Creative Technologies. University of Bolton. Deane Road, Bolton BL3 5AB UK; Senior Research Fellow. University of Bolton. Deane Road, Bolton BL3 5AB UK

Abstract: Relationships between people who meet in virtual worlds are common and these relationships can be long term, in some cases lasting a life-time. Although relationships formed in virtual worlds have invited a lot of recent interest, surprisingly little work has been done on developing computer agents and non-player characters that can actively participate in such relationships. The focus of this review is relational agents, agents that can build long term socio-emotional relationships with users. In virtual worlds, such agents are just starting to emerge; they are more common in other environments but remain few and far between. This review critically assesses the progress of relational agent development and research since their inception in 2005, proposes new areas of research and considers the potential for their exploitation in virtual worlds.

INTRODUCTION

The concept of a computer character that users relate to, and build a relationship with as if they were human, is far from novel. Science fiction has indulged in this idea for decades creating now famous personalities like Marvin the paranoid android [1], Holly from Red Dwarf [2] and the emergency medical hologram in Star Trek Voyager [3]. In Human Computer Interaction (HCI) research, the term social agent is used to describe computer artefacts that use human relationship-building techniques to build a socio-emotional relationship with a user. Like the characters found in science fiction, social agents display idiosyncrasies and personalities that humanise them; the types of personality traits that ostensibly distinguish humans from machines. Significant progress has been made in examining the impact of such attributes as small talk, story telling and humour [4], empathy, encouragement and praise [5,6], hand gestures [7,8] and there are many others. All such research is contributing to the creation of more humanised agents that users can relate to. This review however is focused on a particular subset of social agents known as relational agents. Like all social agents, relational agents are designed to enable a socio-emotional relationship with users. What makes relational agents distinct is their ability to incrementally build a relationship over time by recalling and referring to previous interactions. Relational agents are “computational artefacts designed to build long-term, socio-emotional relationships with their users” [9]. They not only have a memory, they remember specific previous interactions, trivial or otherwise, with the intention of referring to them later so as to evolve relationships. This is an approach that imitates the evolution of conversations as people get to know, love and trust each other.

This review of relational agent research is intended to be comprehensive. Accordingly, a systematic approach was taken to discover published work that will provide an accurate picture of relational agent research from its inception in 2005 to the end of 2008. Relational agents are to some extent an obvious idea but this does not mean that they are common. Indeed very few exist and research relating to them is sparse exposing many areas of empirical deficit. Since 2005, their evolution has been slow and, at the end of 2008, they remain still embryonic. In many areas such as robotics, computer games and various virtual worlds, agents are quickly evolving and many agents are starting to display relationship-building characteristics. It is reasonable to suggest that many such characters in the medium term will inherit the ability to evolve long term relationships with users, effectively becoming characters that ‘know you’. In preparation for this expected mainstreaming of relational agents, much more research is required from sociological, technical and psychological standpoints.

Three sections are presented in this review. Firstly, all existing relational agent application areas are identified and briefly discussed. Secondly the ability of relational agents to build long term socio-emotional relationships is debated, the ability of relational agents to perform tasks normally orchestrated by people is described and related areas of empirical deficit are identified. A final section proposes speculative evolutions that might be incorporated in to the future research and design of relational agents.

APPLICATIONS OF RELATIONAL AGENTS

This first section of the review comprises an overview of existing experimental and practical applications of relational agents. The major thematic areas of application are identified as health and behavioural change applications, leisure and domestic applications and applications related to other areas such as sales, marketing and education. In each area, specific relational agent systems are described and briefly discussed as to their purpose and operation.
Relational agents have generally been implemented as computer generated embodied conversational agents (ECA). They have been configured to operate on client PCs [such as 9-14], large special screens [8,10,15-18 for example], internet sites [19], mobile computer screens [15] and PDA’s [20-22]. Other examples are robots with a humanised interface [23-26]. ECA’s and robots enable emulated face-to-face communication but debatably lack something in that an agent with the appearance of a cartoon character or a robot is obviously not a real human.

The cartoon-like format of relational agents is to some extent a technological limitation. Grolleman et al. [27] however provide an interesting slant on this with a prototype of a character that is not machine-generated but is DVD footage of a real person interacting through a messenger window. Although less like face-to-face communication, a messenger window is closer to the mode that users use to communicate with each other across the internet. The work of Grolleman et al. [27] is still at a prototype stage and although the results of this endeavour are awaited it is interesting to consider this concept in the light of a relational agent (REA) [10] (see below). Subjects communicated with either by telephone or face-to-face. Although the results were not significantly overwhelming, the tendency was for subjects to be more comfortable communicating with her (the relational agent) by telephone. They felt that they knew her better, she seemed more informed, friendly and they felt closer to her. Bickmore [10] speculated that discomfort in communication was a limitation of the ECA; when you see her she is obviously not real while on the telephone this is less obvious. This is discussed further at the end of the review when we discuss the concept of disguise.

Health and Behavioural Change Applications

Relatively few relational agents have been developed and the majority of these can be found in applications related to health and behavioural change. One such example is Laura, the first relational agent (2005). Implemented as part of the MIT FitTrack system she was a computer generated exercise advisor to whom users reported the extent of their exercise [9-14]. Laura’s purpose was to encourage users to take exercise, interact with them on a daily basis, remember previous interactions, participate in small talk and respond with apparent empathy to user activity.

A highly imaginative, virtual world example is the ‘Gay Cruise’ application [19] which encourages the use of condoms among promiscuous homosexual men. Users choose a ‘purser’ (their relational agent) who guides them through various scenarios on a virtual cruise ship instructing the user on HIV prevention. This is an automation of a health promotion technique known as Intervention Mapping [29,30]. The details of Intervention Mapping are not important to this review, but it is significant that due to the purser being a relational agent the whole process, with relationship building as a critical component, can be orchestrated in a virtual world by a computerised agent instead of a human being. A non-relational computer agent would not have been able to do this. This concept is worthy of further consideration; if agents are created that can build long term relationships with users, procedures normally orchestrated by people (in which a certain level of interpersonal relationship is considered critical) will become eligible for automation. Similar plans are underway to create a ‘stop smoking’ coach [27], once again, a relational agent orchestrating Intervention Mapping. Other examples of behavioural change procedures orchestrated by relational agents are Motivational Interviewing [31,32] and the Transtheoretical Model [33,32].

In other-health based examples, relational agents are employed to serve as counsellors who comfort and empathise with people undergoing medical interventions [32], help explain complicated medical documents to people who are not ‘health literate’ [15], teach social skills to children [34], and proactively interrupt a subject’s day by reminding them to rest, to take medication or to exercise [20-22].

Leisure and Domestic Applications

The example of a relational agent with arguably the highest ‘cuteness’ factor is PaPeRo [25]. This is a pet robot that remembers up to 10 faces and displays excitement when it recognises someone. It tells the weather forecast through a Wi-Fi internet link, dances when music is played and blushes when kissed. It also obeys a range of varied commands and displays emotions such as embarrassment, affection and delight. It even attempts to proactively adapt it own personality to be like its owner, introducing such traits as worrying about its appearance. PaPeRo is one example of a ‘sociable robot’ that remembers people, other examples are NeCoRo [26], a robot cat that likewise seeks to build a relationship with its owner and Zeno who is an animated character ‘brought to life’ by Hanson Robotics [35]. To better conceptualise sociable robots, the reader is guided to enter some of these names into YouTube for some quite remarkable footage. In yet another behavioural change example, MIT have now developed a sociable robot [24] that lives with the patient and employs relational agent techniques to provide long term social support for obese people undergoing weight loss programmes [36,37].

The i-cat [23] and major-domo [16] are ‘ambient intelligence’ examples of relational agents. (Ambient intelligence is embedded technology that surrounds users in their home [39]). The i-cat and major-domo are agents that interface to household technology and perform actions like ‘putting on’ some music, turning the TV off, checking for e-
as it relates to the game play. In a fighting game for example, however, has been recording information about interactions that faction to treat the player differently” [41]. In all such cases a range of values are stored about a user or NPC which can be affected by the user’s actions and then referenced by NPCs. The range of values stored can be complex as can the hosting data structures [for example 42] which, in some cases, employ artificial intelligence (AI). The norm in computer games, however, is for user actions to affect a set of values that will later be referenced by NPCs. Specific activities and conversations typically are not remembered in their own right; there is no express purpose of relationship-building over the long term.

The many techniques used in computer games demonstrate the vast spectrum of relationship-related memory techniques that agents in any environment can use. This spectrum ranges from a single number that dictates an agent’s ‘affection’ for a user, to a full database or AI engine that exists for the purpose of recalling previous interactions and evolving a relationship. At exactly which point an agent becomes a ‘relational agent’ is not precisely defined and as agents everywhere evolve towards the existing relational agent definitions the lack of clarity around this boundary will be more obvious. A current example of this is AlphaWolf [34]; this educational game proposes to teach social skills to children. Focused on cubs finding their way within a pack of wolves, it presents an advanced computational model of social relationship formation. What is not clear from the publication is the extent to which this progression is over successive days or logins.

As it stands today the majority of computer game examples cannot be described as relational agents. It is a reasonable observation, however, that findings related to social agents in any HCI context are analogous to those operating in computer games. This underlines one of this articles key propositions, that relational agents are starting to emerge in a range of virtual worlds and many existing agents are likely to be developed as relational agents.

Sales, Marketing and Education

Outside of the application areas mentioned above, this review found only two relational agents. In one example, a relational agent has been used in an electronic learning environment [43] designed for children between the ages of 12 and 15. Then, finally, there is REA [8,10], a life-sized, virtual estate agent. Operating on a very large screen, she provides users with a guide around virtual homes.

AFFECTIVENESS AND EFFECTIVENESS OF RELATIONAL AGENT INTERACTION

Having described and considered the main areas of application of relational agents to practical problems, we...
now turn our attention to studies of relational agents that attempt to answer questions as to the effectiveness of such agents and their level of affect. In this section three questions will be addressed: can relationships be built between agents and real people; how effective are relational agents when given a role to perform; and how do people respond to relational agents. The first of these three questions is the most fundamental; if relationships cannot be built then relational agents are, to some degree, pointless. As will be seen, it is this first question that also leaves the most room for further research.

**Relationship Building Qualities**

The relational agents used in behavioural change applications were designed to be persuasive and it was hoped that they would persuade users to adopt healthier lifestyles. From the outset, this became questionable. In the first example [9,12], Laura attempted to encourage exercise but when operating with 101 young people, 69% of whom were MIT students, she was ineffective when compared to the control; the relational agent’s subjects did not do more exercise. Despite this failing, significant positive findings came from this research in that a substantial majority of subjects ‘liked’ the agent and wanted to carry on working with her as compared to the non-relational control. Although some people did not share this experience, unable to conceptualise the idea that ‘she’ was anything more than an algorithm, the majority of participants felt that they did have a relationship with Laura and they even felt cared for [11]. These results were echoed in a similar study [10] that also operated with young people (91 participants with a mean age of 25). Although the majority of subjects warmed to the character there were a few who did not, providing such responses as ‘Laura is not a real person, and therefore I have no relation whatsoever with her’ [10]. Overall, this suggests that for the majority of people, the relational qualities of the agent made the application significantly more attractive and enabled some kind of bond (a relationship) to be created between the user and the character.

In these early experiments, many subjects reported an annoyance with the relational aspects of the character (small talk, facial expressions, for example) after the initial novelty period wore off, primarily due to its repetitive nature. It is speculated [13] that this was more of a limitation of the Laura character than the relational agent concept. Encountering a real person who insisted on similar small talk at each interaction would also be highly tedious and aggravating. It was further speculated that this annoyance would deepen over time but in a rather unusual twist this proved to be false; in a later and longer experiment [13,14], the relationship actually deepened over time. In this later experiment, working with older adults (average age 74), Laura won the affection of her users quite remarkably. On a scale ranging from stranger (1) to close friend (7) the mean came to 6.8. These results are reflected in the comments that the users made about the system being nearly all positive and demonstrating elements of a human-like relationship with the character. These experiments also proved to be persuasive with the relational agent subjects doing much more exercise than the control group. Although participation with the system decreased from an average of 4.6 to 2.3 contacts a week at the end of the 60 day trial, participants indicated they would like to continue using the system with an average weighting of 6.4 (1=“not at all” and 7=“very much”). These results are significant and strongly support the arguments that relationships can be built and that relational agents make applications more attractive to use. The other initial experiments [9,12] also supported the relational agent argument but there is no question as to the greater success of this later, longer experiment.

The key difference could be the age and social mix of the participants. In the more effective application, the adults were older (average age 74), nine out of the ten scored a low reading literacy, all were female, half had never used a computer before and they were nearly all (90%) African Americans. A considerably different group to the young educated people who participated in earlier trials. Any of these differences could be responsible. Consider for example the ethnic mix – one participant reported they liked the fact Laura was a “person of color” [13]. Various studies have suggested that automatic social response and stereotyping are as applicable to human computer relationships as normal communication between humans [44,45] and that communications from agents of the same ethnicity as the user are perceived to be more attractive, trustworthy, persuasive and intelligent [46] – all important components of human relationships. Interestingly, a number of applications now enable the user to choose their own relational agent [19,47] and include the option to select an agent they are naturally attracted to in some way.

Personality types were found to modify the response that users had to REA the multimodal virtual estate agent [8,10]. Extroverts and introverts reacted differently as did passive and active subjects. Extroverts for example trusted REA more in embodied interactions but, over the telephone, being an extrovert had no impact. In face-to-face contact, introverts trusted the purely task-orientated version more while extroverts trusted the agent significantly less in this mode. Passive subjects felt more comfortable interacting with REA than active subjects did (regardless of the communication mode) while preferring face to face communication to communication over a telephone. It was hoped that such personality experiments would help to diagnose the factors that dictate if an agent will be trusted and/or a relationship built but more recent work indicates that such responses are not standard with all relational agents reflecting instead the perceived personality of the agent [48]. REA, for example, acts like an introvert and if she were real this would invoke various reactions from various personality types. These findings propose one of two conclusions; either natural inter-personality reactions transfer to the world of relational agents or different personality types react to agents differently. Although there could be elements of truth in both conclusions, analysis in this case suggests the former.

When a relational agent was developed for an electronic learning environment [43] designed for children between the ages of 12 and 15 years, the results were mixed with a marginal majority preferring the relational agent to the control. When summarising responses, Gulz is clear that many liked the relational agent while others viewed the small talk, for example, as annoying and unnecessary [43]. This somewhat contrasts with the findings related to Laura
and REA [9-14] as Gulz points out. It has to be concluded that the success of relational agents in establishing relationships with people is varied and insufficient work has been done to diagnose the critical success factors. It has already been suggested that this could relate to the socio-economic groups involved, the similarity between the agent and the user, personality types or the IT literacy of the user. Looking at this educational example [43], it might also be suggested that age is a factor. Although these experiments employed different agents and quite different scenarios, a trend can be observed. Gulz, when working with children found a 50/50 split in the establishment of a relationship with the character. Operating with young adults, Bickmore [9,10,12] found that the majority were able to establish a relationship with the agent, but when working with older adults, 100% of participants established a relationship [13].

**Human Response and Agent Capability**

It has already been observed that relational agents are capable of performing tasks designed for human orchestration [19,27,32] and that when it comes to persuasion [9,12] their effectiveness as compared to the control is dubious. In another example, where an agent was used to explain complicated medical documents to the ‘non-health literate’ [15], the effectiveness was once again questioned and the relational agent’s subjects did not understand the literature any better than when they interacted with the control, which in this case was a real person following a similar script to the agent. In these examples it should be clarified that on no occasion were the relational agents found to be inferior to a control. However it would be a mistake, without further evidence, to conclude that relational agents are therefore more effective at performing tasks than people or non-relational agents.

As previously discussed, despite not improving the outcome of a procedure, relational agents do make a number of new activities eligible for automation. It is also clear that relational agents attract user participation. In the case of agents being employed to explain complicated medical documents [15], the subjects actually preferred the agent to dealing with a real clinician (which, as mentioned above, does not mean that the agent was necessarily more effective in the task of conveying information than the clinician). One apparent reason for this is a perception that clinicians are intelligent and busy people and that a patient has limited time with them and does not want to display ignorance and the consequent loss of face. With the relational agent, time is not limited and ‘face’ is not at risk. Another example supports this argument, when a relational agent was employed as a museum guide [17,18] a significantly smaller majority of respondents said they would prefer to interact with the agent than a real person, perhaps because the public perception of museum guides is typically less deferential than clinicians. It is reasonable to surmise therefore that there is a correlation between how important a person is and how much people will welcome a relational agent imitation of them performing aspects of their role. It is also significant that in the case of the museum guide, although the majority of people who preferred a relational agent over a human was smaller than the clinician case, this was still a majority (56% preferred the agent to a real person, 31% said they would prefer a person and 13% were unsure). It appears that the most people simply ‘like’ interacting with relational agents, certainly more than interacting with other agents and in some cases they are preferred to real people.

**DISCUSSION**

As this review has shown, relational agents have only been exploited in a few application areas. To summarise, relational agents have been used to orchestrate behavioural change procedures [9-14,19,27,32,36,37], to explain documents [15], to counsel patients [32], as museum guides [17,18], as robotic pets [25,26], in ambient technology [16,23], in e-learning [43]; in sales [8,10] and, to a very limited extent, in computer games [34]. This is a very small cross section of application areas and in most areas there are only one or two examples. These existing cases do however prove the concept suggesting that in the future, relational agents could be applied to any number of applications.

It has also been shown that relational agents can be used to perform procedures created for human orchestration for which a progressive interpersonal relationship is required. Only three examples exist, namely ‘Motivational Interviewing’ [32], ‘Intervention Mapping’ [19,27] and ‘the Transtheoretical Model’ [32]. It is proposed here that many relationship-dependent procedures which currently demand human orchestration are now eligible for automation. Some health intervention procedures are among those that have already been investigated but many similar procedures exist in the world of business administration, pedagogy/andragogy, sales, religious/cultural practice and leisure to name a few. It is not by any means proposed that all such activities could be orchestrated by relational agents but there are a substantial number that could be. In the immediate future, virtual worlds could provide a range of automated services currently offered only by real people. Given the current application areas, it is clear that these could involve various forms of counselling, mentoring and coaching. Perhaps a relational agent could act as a friendly guide to those new to virtual worlds.

Thus, we may even suggest that even without the specific objective of orchestrating a procedure, relational agents could improve a virtual world experience in other ways. Relational agents have been shown to make computer interfaces more attractive to users by encouraging them to return to the application [9,11-14] and there is no reason why this ability should not transfer to virtual worlds. In leisure environments for example, if NPC’s were able to recall previous encounters this would potentially make the environment more realistic and attractive due to the evolving relationships. The trick would be creating agents with a realistic (not perfect) memory for salient events/conversations or repetitive activity reflecting the manner in which real people remember those they interact with. Remembering 100% of interactions, trivial or otherwise, is what computers do, it is not what people do and it is not how relationships are built.

Evidence presented in this review suggests that people of different ages, genders and ethnic groups respond differently to the same relational agent and, indeed, that the relationship may be improved by matching, for example, racial characteristics of the agent to the user. It would be a reasonable assumption to suppose that such relationships
may be further strengthened by matching yet more factors such as gender, age, race and culture creating relational agents that reflect the type of person to whom the user would naturally affiliate or be attracted. Perhaps back stories for the agent (in effect, a simulacrum of a lived life of experience) might be developed to provide a social and cultural milieu for the user to recognise and respond to.

Some participants in relational agent experiments refused to accept that a relationship with an agent is possible on account that they are ‘not real’. It has already been proposed that a cause of this could be the inability of technology to create a human-like agent. There are other factors potentially responsible though. A relational agent, for example, did not go to primary school, it does not fear death, fall ill, have a mother in law, get tired, have children and so on. Almost by definition the user has nothing in common with the agent. An avatar could represent someone who has a genuine interest in the success of a procedure, someone in authority or a person to whom the user might naturally be drawn (same age group, level of education, for instance). This contextualisation (that is, the representation of a familiar figure with all his/her quirks, characteristics and, perhaps, a photo-realistic use of the person’s face and body) of the agent could also provide the grist for the agent’s idiosyncrasies, its moods, bad jokes and nasty habits (humans are not perfect and neither should be relational agents).

It is not yet known precisely what the factors are that induce a relationship between agent and human. Responses to relational agents tend to be polarised between the development of some form of relationship and the outright refusal to accept such a possibility. We have already discussed potentially significant factors affecting this and so now turn our attention to other persuasive techniques. It may be that one technique lies in disguise given that humans are adept at recognising, and therefore filtering by, difference. This facility is one of the factors behind the ‘uncanny valley’ phenomenon [49] experienced when users are presented with a visual computer character purporting to be human – although technology has made great strides in this area, users are still able to recognise the lack of subtle human facial characteristics or certain minute discrepancies in lip synchronisation. In this scenario, there is the added factor that the character is all too clearly computerised (as betrayed by its context). Disguise, the deliberate removal of at least one mode of communication, might take many forms. The simplest and perhaps most effective might be to remove the visual embodiment altogether and to pursue the relationship mediated by other communication tools such as telephone (such technology already exerting a distorting, and therefore disguising, effect on the voice), e-mail and other forms of textual digital communication. As well as exploiting disguise, this would more naturally imitate one of the ways in which modern humans regularly communicate.

It has already been mentioned that computer games are likely to make increasing use of relational agents and here too the concept of disguise may be a significant factor. In many online games, human players are represented on screen to other human players through the use of an avatar and this avatar is the same visual character as might be used by an NPC in the same game. Visually, there is no distinction and players in such games can build long-term relationships with other players having only ‘seen’ them through the guise of their avatars. Relationships are built upon a shared team ethos, for example, or through in-game voice and chat messages. This visual disguise is one that computer game relational agents may adopt allowing game developers to concentrate on human-like behaviour instead.

In an abstraction from reality, most relational agents are contained within virtual environments and will only be encountered when the user chooses to ‘visit them’. Exceptions to this rule are the robotic examples [23-26] and those that were designed to reside on a PDA with the express purpose of interrupting a subject’s day [20-22]. In the real world, most people have experienced relationships in which the second party never initiates contact. This can become frustrating and, in most cases, the first party will also decide to let the relationship die. This obvious aspect of relationship building has not been transferred to the virtual world; apart from the few exceptions where the nature of the hardware facilitates it, agents do not proactively contact users. It has already been suggested that an agent could be developed that uses textual digital communication in an attempt to imitate comprehensively the manner in which people communicate through technology. Such an agent could also be programmed with relative ease to proactively make contact with the user, encouraging the relationship and reducing the boundary that separates real and virtual worlds. The suggestion is that proactive communication would make the agent more realistic and would improve its ability to build relationships.

To be involved in relational agent research at this juncture is to realize that it is a field of research at the frontiers of knowledge and that much, much more remains to be done. It needs recognition that research in the area requires a multi-disciplinary approach covering, for example, areas such as artificial intelligence, computer science, animation, speech synthesis, psychology, cognitive science and behavioural science. It is our hope that this review will prove beneficial to those wishing to participate in this exciting field of knowledge not only by summarizing the current state of the research but also by providing pointers to new opportunities for enquiry.

REFERENCES


