CHAPTER SEVEN

Pleasure versus Efficiency in User Interfaces: Towards an Involvement Framework

ANTONELLA DE ANGELI, PAULA LYNCH and

GRAHAM I. JOHNSON

Advanced Technology and Research, NCR - Self Service Strategic Solutions, Kingsway West, Dundee, DD2 3XX, UK

7.1 BEYOND USABILITY

The concept of usability is regarded by many as a milestone in the history of computer system design, specifically the design and evaluation of user interfaces. Since the term entered common usage in the early 1980's, the consideration of usability has greatly impacted on the way in which many interactive systems are developed. Usability compels designers to think from the very beginning about end-users. Therefore, it has contributed to the evolution, from the traditional top-down design approach (all requirements were specified in the planning phase and then developed by stepwise refinements) to a more iterative-design approach (evaluation and implementation are closely linked, user input at key stages, and ongoing requirements definition and system specification).

Despite the apparent current popularity of usability thinking, the acceptance of the user-centred metric has not been easy in a development world dominated by a previously unquestioned system-centred philosophy. Historically, the latter approach assumed that users could, and would, adapt to whatever was built. Training, support documentation and 'Help' functions were considered to be the most appropriate solution to serious interaction difficulties. Only within the usability framework has the end-user become the focus of the design process. The recognition that usability is a fundamental aspect of product quality as well as of marketing is a fairly recent reality. Nowadays, much usability research effort is devoted to the development and validation of cost-effective evaluation tools to encourage the integration of usability issues into design (e.g. Nielsen 1993; Johnson 1996; Jordan *et al.* 1996; De Angeli *et al.* 2000).

Currently, the value of the usability framework is so well acknowledged (ISO 1991) that proposing major revisions will no doubt sound outlandish to many usability and human factors specialists. A number of objections can be raised to the modification of a successful metric, which is reliable and capable of driving the design of effective systems. This chapter attempts to meet these objections, proposing an evolutionary perspective.

Our belief is that the need for re-examination and update of the general usability framework is prompted by the powerful combination of technological progress, consumer expectations and the evolution of novel applications. A new generation of interactive systems, utilising varied animated characters, sophisticated behaviours, embodied conversational agents and interfaces with 'personality', will revolutionise the way people interact with computers. enlarging the bandwidth of communication to fully include social and affective dimensions (Picard 1997; Cassell 2000; Cassell *et al.* 2000). These next generation systems will elicit different psychological reactions from users; hence, they need to be evaluated in a specific framework respecting their key characteristics.

Human-Computer Interaction (HCI) has many examples of successful frameworks and approaches that have been significantly modified as a result of the movement of the state of the art of technology. Consider, for instance, the GOMS model, a cognitive modelling tool to represent the procedural knowledge that a user must have in order to carry out specified tasks on a system (Card, Moran and Newell 1983). GOMS analyses apply well to situations in which users perform simple and linear goal-oriented tasks which they have previously mastered. Therefore, it is well suited to model text-editing tasks, but has difficulties in extending towards creative tasks, involving parallel activity, such as those supported by contemporary multimedia systems.

HCI is a dynamic discipline whose object of study is continuously evolving (De Angeli 1997). The interrelation between HCI and technology can be represented as a loop: HCI modifies technology, the new technology modifies HCI. Computers change, users change: As a result, the interaction is different and need to be investigated with different research tools. A constant update of the theoretical and methodological frameworks within HCI is required, as a consequence of this evolution of technology.

The chapter discusses our vision of the future with respect to personality-enriched interface developments, providing some initial thoughts on an involvement framework, and augmentation of a traditional usability engineering perspective. A brief review of the pertinent literature and the results of a preliminary survey of an exemplar financial interface embracing critical attributes of personality are presented in the following sections of this paper.

7.2 SOCIAL ARTIFACTS AND RELATIONSHIP TECHNOLOGIES

In recent years, the interaction bandwidth between users and computer-based systems has exponentially increased and the future is promising even more stimulating scenarios. Also, digital networks have moved far away from their original habitat: they have moved from academic and industrial settings to private and public spaces. In parallel, the Internet has evolved and the consensus is that the future impact of the Internet on all personal and business relationships will be tremendous.

In the Network Economy (Kelly 1998) the real business is in building relationships between customers and information providers. According to the Knowledge Lab vision (Lynch, Emmott and Johnson 1999), the future of interactive technologies is not centred on raw information or on Information Technologies (IT). Rather, the focus will be on *Relationship Technologies*. These are primarily aimed at gaining the attention of consumers, attracting them, understanding their needs, communicating with them, supporting them, and inducing them to use the services. The term was originally proposed by the NCR Knowledge Lab, to define existing and emerging technologies and models that enable, support and enhance relationships between customers and providers, or between groups of customers. Here, the definition is extended to include those direct relationships between users and (social) interfaces.

Historically, IT has been aimed at creating effective machines, encompassing all of

the rationality, logic and abstract knowledge available to human beings. IT has typically produced cognitive artifacts: objects that store, manipulate and retrieve information (Norman 1991). Cognitive artifacts are artificial tools made by human beings to support representational functions. In contrast, relationship technologies are going far beyond efficiency, building social artifacts – agents that create and maintain meaningful relationships between users, groups of users and interfaces. These systems will massively impact on the focus of and approaches within HCI, introducing emotions, affect, attitudes and social intelligence, as well as enjoyment, pleasure and humour in the interaction with computers and via interfaces.

From the user's point of view the difference is enormous. Cognitive artifacts mainly support instrumental activities, from perception to problem solving or reasoning. These artifacts elaborate commands according to rational rules and are, therefore, predominantly passive objects, which can be completely under our control. A linear causal model can explain the interaction between them and users. On the other hand, to build lasting and meaningful relationships, social artifacts need to be active agents, capable of responding socially to users, appealing to emotional states, dynamic, showing a 'sense' of personality and attitude. These artifacts have the potential to enhance the anthropomorphic perception elicited by computers: they can easily be perceived as animate, aware, and active (De Angeli *et al.* 1999).

Little published research exists on whether the intentional introduction of social dimensions and emotions in HCI is a positive (or a negative). Until very recently, the prevailing notion was that computer systems must reduce learning times and task-performance times, as well as error rates. Peripheral attributes, such as playfulness or humour, were typically considered wasteful – something which could distract a user and cause them to take their work less seriously. Nevertheless, a new trend is emerging in both the Artificial Intelligence (AI) and HCI community (see e.g. Proceedings of UM'99 Workshop on Attitude, Personality and Emotions in User-Adapted Interaction; Proceedings of the 3rd I^3 Annual Conference, Workshop on Affect in Interactions 1999). Researchers are starting to examine issues of personality (e.g. Lynch *et al.* 1999) and humour (e.g. Morkes, Kernal and Nass 1999) in interfaces and interaction. The latter team has demonstrated that humour can have similar effects, whether the source is a person or a computer system. In principle, humour can be used to improve the overall user experience, helping to relieve the tension that is often associated with the performance of a demanding task.

The following section of the paper describes how amusement and enjoyment can be introduced in a stereotypically serious context – that of a public self-service cash machine, an ATM (Automated Teller Machine). In this application, the purpose is to service consumer banking needs, normally considered a chore, and rarely (if ever) regarded as a pleasurable or enjoyable experience.

7.3 USER INTERFACES FOR BANKING

Background

Many of today's financial consumers visit their (physical) bank branch less than three times a year. For these consumers, interaction with their bank normally takes place remotely via technology (ATMs, telephones, and personal computers). Financial service providers recognise the intrinsic importance of user interfaces as their primary means of contact or relationship with customers, and are striving to further enhance the consumer

experience, as a way to differentiate from competitors, and as a means of improving service quality. In recent years, research work in this area of financial user interfaces has focussed on new interaction scenarios and interfaces (e.g. Johnson 1995; Johnson 1996b; Johnson and Westwater 1996; Baber, Johnson and Cleaver 1997; Coventry and Johnson 1999). Research has also investigated evaluation and design methods for usability in self-service interface solutions (Johnson 1993; Johnson and Briggs 1994; Westwater and Johnson 1995).

The NCR Knowledge Lab is exploring many opportunities to make consumer experiences of e-commerce applications more engaging, personalised, and relevant (Lynch *et al.* 1999). We are investigating how interaction with financial user interfaces can be improved by introducing a sense of personality, attitude or character into the equation. As a result, we created a project to examine the potential of, and reactions to, a prototype (ATM) user interface which is enriched by a specific (type of) character: we affectionately called her 'Granny'.

'Granny': a financial interface with 'personality'

The following scenario describes how a virtual user interface (see Figure 7.1) could be employed, within the common context of an ATM user interface, to facilitate consumer interaction with a financial service provider in the near future:

Emma, a 22-year-old maths student visits an ATM, close to where she lives. A few weeks ago, she chose 'Granny' as her personal banking interface. There were many other 'agents' or personalities that she could have chosen; however, Granny looked like she could be a lot of fun. Once logged on, via a PIN, Granny greets Emma and displays her personalised options such as cash (withdrawal), her accounts and local information.

The overall layout, animations and icons on the screen reflect the general character of Granny – for example, clicking on the 'handbag' icon Emma can access her diary and shopping list. Granny is learning about Emma's preferences by overtly observing her transaction behaviour and by retrieving relevant personal details from the (financial provider's) data warehouse. When Emma requests £30 Granny reminds Emma that it's Bob's birthday next week. In the course of the transaction, Granny prints a receipt presenting Emma with her daily horoscope.

Granny has had a very brief life thus far, and has moved from a basic concept to a full-blown (interface) application within an ATM, which can dispense real cash very quickly. The concept came from a brainstorm session, which concluded with a role-play, two of the authors 'acting' the part of Emma and Granny. Sketched storyboards and then an interactive prototype swiftly followed this stage. The move from PC to a touch-screen ATM represented the final stage of her development.

Granny is an example of a social artifact that attempts to inject a specific personality or character into the interaction between a financial service provider and a consumer. In this context, personality is defined as a stable set of traits that determines the artifact interaction style, describes its character, and allows the end-user to (understand and) predict the artifact's general behaviour.

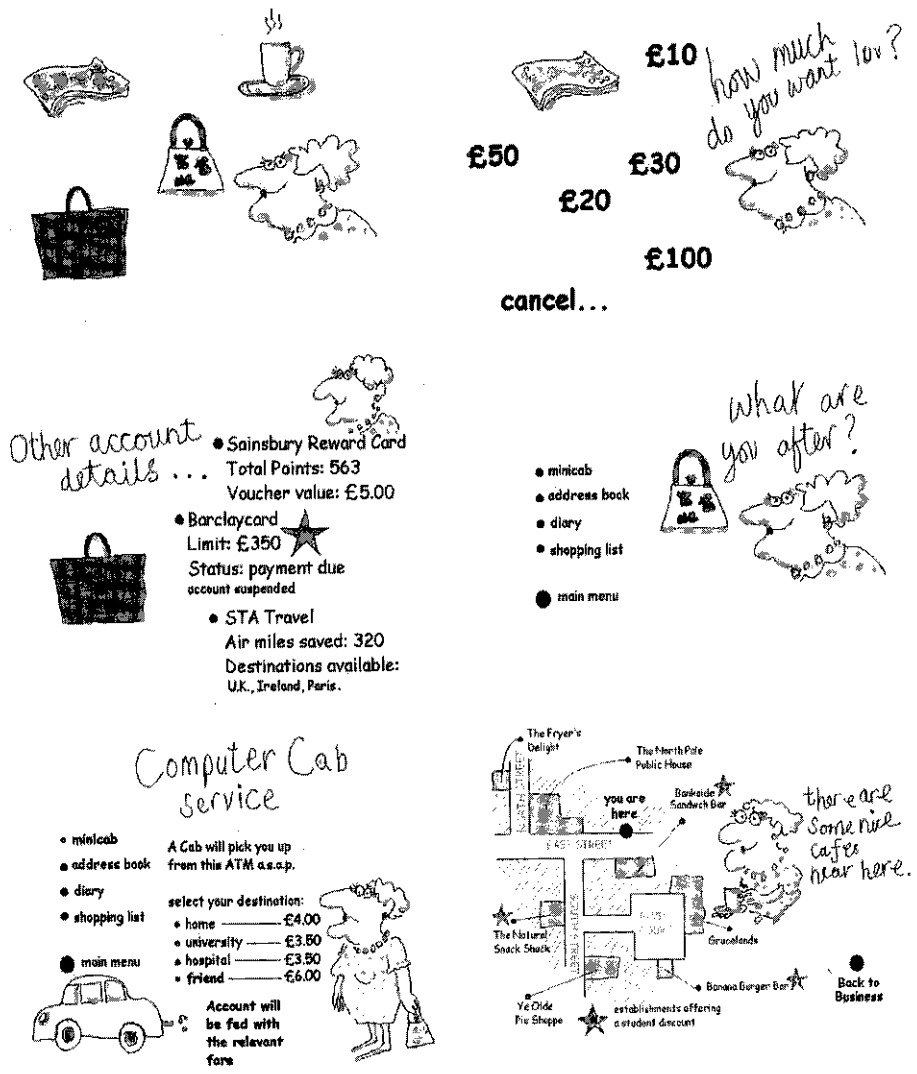

Figure 7.1 'Granny': A financial user interface with personality – example screens

7.4 EVALUATING SOCIAL ARTIFACTS

Despite the general importance attributed to the concept of personality in the development of intelligent (user interface) agents, little is known yet about what type(s) of personality should be implemented. The prevailing approach, typified by Granny, is deliberately anthropomorphic in nature, but more research is needed before effective synthetic personalities can be developed. We have begun to investigate the potential for personality profiles to be used for e-commerce social artifacts, specifically synthetic characters. Starting with the five-dimension OCEAN model of personality (see e.g. Costa and

McCrae 1992), we outlined the possibilities using the five dimensions: Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness. Each dimension contains a set of traits that tend to occur together and can be conceptualised as a continuum ranging from two opposite personalities. We are working towards a characterisation using this framework, based upon these five factors, the context and task/application in hand. It strikes us that one of the biggest challenges is in developing a framework that can *reliably* measure the effects that social artifacts have on consumer behaviour and attitudes. For example, how effective is Granny in building strong and lasting (versus amusing, casual, superficial) relationships with consumers.

ISO Standard 9241 defines usability as 'the extent to which a product can be used by specified users to achieve specified goals with *effectiveness*, *efficiency* and *satisfaction* in a specified context of use'. *Effectiveness* refers to the accuracy and completeness with which specified users achieve specified goals in particular environments. *Efficiency* refers to the resources expended in relation to the accuracy and completeness of goals achieved. *Satisfaction* refers to the comfort and the acceptability of the system for its users and other people affected by its use. Such a framework refers to computers as tools. Despite being a multidimensional concept, usability mainly relies on objective parameters describing task performance. Indicators of effectiveness include quality of solution and error rates; indicators of efficiency include task completion time and learning time. Satisfaction is the only parameter that directly addresses feelings, opinions and attitudes of the user. Nevertheless, this dimension has traditionally been treated as a measure of perceived effectiveness or efficiency. This approach, explicitly or implicitly, relies on the assumption that users are satisfied when they perform their task efficiently. Nevertheless, the correlation between the three major attributes of usability is still not fully understood, and in the case of complex tasks has been proved to be quite weak (Frøkjær, Hertzum and Hornbæk 2000). These findings breach the traditional framework suggesting that users could be satisfied by something other than efficiency.

Cognitive artifacts are designed to help perform tasks effectively: They need to be simple and efficient. Social artifacts are designed to establish relationships; they need, at some level, to be pleasurable and engaging. The major aim of social artifacts is to create and maintain relationships. For all these reasons, social artifacts do not just need to be *user-friendly*, they need to be *friendly to users,* responding socially to them and showing a 'sense' of personality, attitude and understanding.

Applying the media equation

Following the media equation paradigm, as stated by Reeves and Nass (1996), the framework to evaluate social artifacts could be directly imported by social psychology or by the psychology of personality. According to the media equation (*media = real life*) individuals' interactions with computers, television and new media are fundamentally social and natural, just like interactions in real life. This assumption implies that the same social rules guiding human-human communication are equally applied to HCI, even though such behaviour is not necessarily conscious. Research has demonstrated that people engage in polite, socially desirable behaviour when interacting with computers (Nass, Moon and Carney 1999). Moreover, users apply gender stereotypes to computers (Nass, Moon and Green 1997) and human personalities (Nass *et al.* 1995; Moon and Nass 1996; Nass and Lee 2000). Also, when a person is monitored, either by an animated agent or a human, the same psychological reactions are elicited (Reeves and Rickenberg 2000).

The media equation implies that computers should be evaluated by the same instruments used by psychologists to measure the strength of interpersonal relationships. Hence, the evaluation framework for social artifacts can be based on personality tests,

attitude and motivation questionnaires. Moreover, the same social rules and norms that drive interpersonal interaction can explain user behaviour. In our opinion, the limit of the media equation can be identified in its generality: the media equation applies to everyone and to all media (De Angeli *et al.* 1999). In this view, social behaviour encompasses any exchange of meaning between users and computers. Such an assumption generates a paradox: every instrumental act (like pushing a key) is also a social act. Taking the media equation to extremes poses the problem of differentiating computers from users, instruments from agents, cognitive from social artifacts. Recently, the same authors have started criticising the initial equation, claiming that HCI and CMC (Computer-Mediated Communication) are not identical (Morkes, Kernal and Nass 1999).

Many empirical studies have demonstrated that face-to-face communication is not an adequate model to explain natural-language interaction (Bernsen, Dybkjær and Dybkjær 1998). Talking to a computer, people maintain a conversational framework, but tend to simplify the syntactic structure and to reduce utterance length, lexicon richness and pronoun usage. We can reasonably expect similar simplification effects affecting social attribution for synthetic characters. The challenge, however, is in understanding which key dimensions are used to evaluate social artifacts and in then translating them into a reliable evaluation framework. The next section describes a preliminary empirical study aimed at understanding how users actually perceive social artifacts, taking Granny as the example.

7.5 A PRELIMINARY STUDY OF 'GRANNY'

Our initial study was aimed at identifying the fundamental perceptions of three different financially related *targets*. These were:

- *Traditional ATM*, representing a typical example of cognitive artifact, an operative tool providing a familiar functionality of basic financial information and dispensing cash.
- *Granny*, as the social artifact prototype: 'she' has a clear personality and is able to adapt to the specific user and context where the interaction takes place. Further, this prototype introduces amusement and humour to the interaction.
- *Cashier* (a human teller), representing a simple control condition, to assess possible differences between social artifacts and human operators.

As our working hypothesis, we assumed that the general perceptions of, and mental representations encouraged by, social artifacts are essentially different from those concerning cognitive artifacts and human operators. Moreover, we expected that the overall reaction towards social artifacts would be generally positive.

Approach

A few days before the survey study, all participants were introduced to Granny during a formal demonstration at the NCR Knowledge Lab, in March of this year. The prototype was installed in a standard multimedia ATM that uses a touch-screen. The interaction started with inserting a bank card and ended with the provision of cash and a personalised receipt. Participants could not directly interact with Granny, but the speaker (demonstrator) showed all the functionality. The demonstration lasted for almost ten minutes, after which participants could ask questions.

Twenty-two Knowledge Lab employees were invited to participate in the study via

an e-mail based survey form. The survey instrument itself was an Excel file, composed of three sheets (one for each target). Participants were asked to list six adjectives or short sentences that described each target. Respondents were explicitly instructed to follow the order of presentation of the sheets, switching to a new one only when the previous one was completely filled. To minimise any carry-over effects in the evaluation, the order of presentation (of the targets within the survey) was counterbalanced across participants.

Coding and analysis

Data were available from a sample of 16 respondents (5 F, 11 M; yrs mean = 29), yielding a corpus of 266 adjectives or short sentences (see Figure 7.2). Of these 90 described the cashier, 93 the ATM and 83 Granny. The corpus was initially cleaned up, deleting all the meaningless sentences (N=10). All of the remaining adjectives were then aggregated in semantic categories following the synonymous taxonomy, provided by WordNet 1.6,
http://www.cogsci.princeton.edu/.

The analysis suggested the existence of three superordinate *Evaluation Dimensions (ED)* along which the corpus could be differentiated. Each ED reflects a particular aspect of the targets, which were described in terms of:

- functional quality,
- aesthetic quality, and
- social quality.

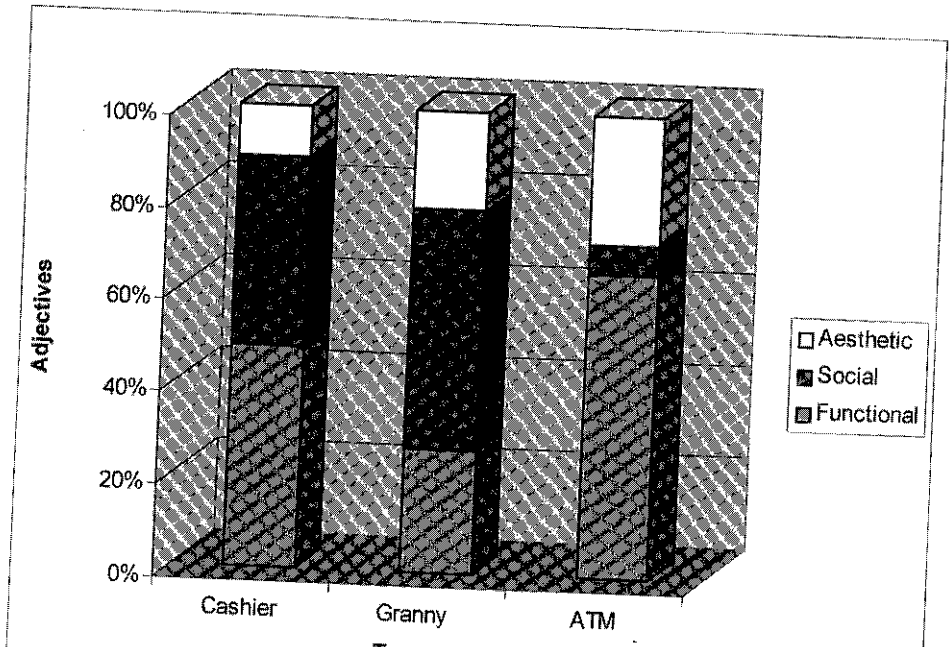

Figure 7.2 ED's

The *functional* ED provides an instrumental description of the target. Here, the ATM, the Cashier or Granny were portrayed as *tools* used by customers to perform financial tasks. This ED mainly reflects a functional perception describing how the target works with respect to the available functionality, the procedures to operate it, overall effectiveness, efficiency or usefulness. The ED category includes all those adjectives describing the target in terms of operative abilities related to problem solving or reasoning, interaction abilities related to the communication of task-related information, accessibility of the service, and novelty of the service. It is evident that the functional ED closely resembles the traditional usability concept: the target is described in terms of a tool, and in relation to the task to be executed. Example adjectives categorised in this ED were: 'reliable', 'service oriented', 'numerical', 'mechanical' and 'efficient'.

The *social* ED reflects a more complex representation of the target. Here, the functional aspect of task execution is disregarded and the stress is on the (personal) relationship between the customer and the target. The category provides a perception or mental representation of the targets as agents capable of actively shaping the relationship. The social ED includes all those adjectives describing personality traits (with the exception of intellectual capabilities related to task execution, previously included in the functional category) motivations, attitudes, feelings or emotions. Example adjectives classified within this dimension include 'chatty', 'engaging', 'personal' and 'smiling'.

The *aesthetic* ED reflects a mental model of the target in terms of its overall physical appearance and perceptual qualities. With this dimension, the stress is on the sensorial perception of the target, primarily on the 'look and feel'. The category includes descriptions of physical aspect, design or form of the target relating to visual, tactile and auditory perception. Example adjectives included 'colourless', 'ugly', 'pink', and 'grey'.

Each adjective or short description contained in the corpus was tabulated in one ED. Double scoring was conducted for 100% of the corpus and any discrepancy was resolved before further analyses were performed. The adjectives were then further classified according to their value into two categories: positive vs. negative quality. The double scoring had a reliability of .89. The adjectives on which a correspondence was missed were not included in the analyses, since their value was considered highly context-dependent (most of which referred to humour or amusement).

Results and discussion

The pattern of ED percentages as a function of the evaluation target is illustrated in Figure 7.2. It clearly emerges that the typology of the target strongly affects the elicited description, the perception or attributions made, and that there are clear differences between the three stimuli of Granny, ATM, and Cashier.

Let us first look at the key differences between the two 'machines' (traditional ATM and Granny). The ATM is mainly described in terms of functional qualities (over 65%) related to task execution, which may not come as too much of a surprise. In contrast, Granny is primarily described in terms of social qualities (over 50% of the adjectives provided for Granny), referring to the relationship between the consumer and the system via Granny. From an evaluation point of view, this difference is not trivial. Granny is a social stimulus; the ATM is mainly a cognitive one. It is fundamental to understand that the two stimuli elicit very different psychological reactions from the end-user or consumer. Social stimuli are more complex than cognitive ones: they are more likely to be two-way interactive agents, perceiving while they are perceived, changing because they are perceived, and involving the observer. Further, many important attributes of social stimuli are not directly observable (e.g. traits, intents, attitudes), and the accuracy of observations is difficult to determine. All these characteristics have to be taken into

account when evaluating the user's reaction to social artifacts and will deeply complicate the evaluation framework.

As regards the comparison between the human cashier and Granny, a counterintuitive result emerged. Note the relative difference between the percentage of functional and social ED (which shows that for the cashier or teller there were 48% and 41% functional and social adjectives respectively). Surprisingly, many more (by relative proportion) social attributes were produced to describe a machine (in the form of Granny) than to describe a human being: 52% for Granny and 40% for the cashier. Such a gap can probably be attributed to a typical anchoring effect affecting the evaluation. The cashier is a *functional human* whose social characteristics are irrelevant if compared to important others, such as friends or relatives. We see the effect of people's perception of the *occupation* of bank cashier, rather than the description of Doris the teller at the local branch. Granny is (intentionally) a *social machine* whose functional characteristics are largely irrelevant if compared to stereotypical machines.

The percentages of positive adjectives in each ED as a function of the evaluation target are presented in Figure 7.3. It is clear that Granny received a very positive overall evaluation. The weakest aspect is the functional dimension. Most of the negative attributes were related to the time demand associated with the use of Granny as a functioning ATM. From a purely functional point of view, the ATM appears to be the best solution (in that, of the three targets, it has the most positive functional adjectives). However, it does not elicit any positive reaction along the other two dimensions.

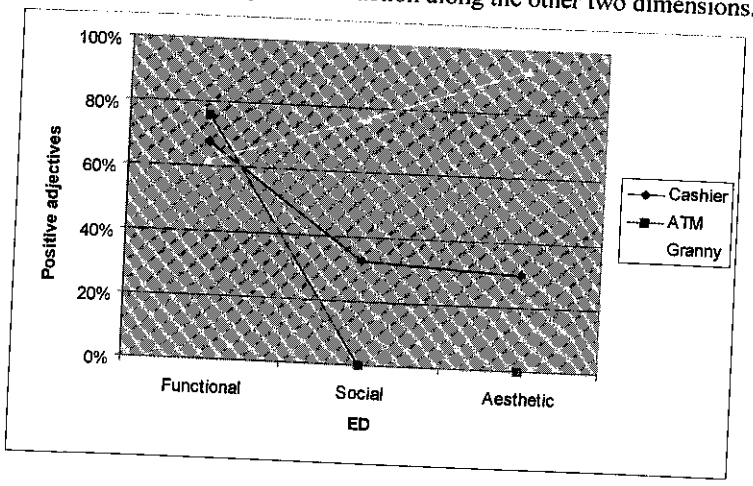

Figure 7.3 Percentage of positive adjectives in each ED as a function of evaluation target

The physical appearance of Granny was highly appreciated. With a very few exceptions, the sample evaluated it as appealing, attractive, and colourful. Also, the social dimension was considered positively, but it can still be improved. Most of the people evaluated Granny as friendly and entertaining, but someone pointed out that it could be over-friendly and silly.

The evaluation trend attributed to the human operator is somehow surprising. Only the functional dimension received a majority of positive evaluations. From a social point of view, human operators were often perceived as unfriendly, bored, depressed or sad. Again, we are most probably witnessing associations to do with the occupation per se with its stereotype of repetitive and mundane counter-based work.

Our initial study was simple in format and our results are preliminary, but they

clearly demonstrate that social artifacts tend to elicit particular representations and attributions, which differentiate them from both cognitive artifacts and people (such as a cashier) performing exactly the same task set. Hence, the results offer some support for our basic premise that both the usability framework of ISO and conventional wisdom and the media equation are not fully adequate to provide a reliable framework for social artifacts evaluation.

7.6 TOWARDS THE INVOLVEMENT FRAMEWORK

Evaluating social artifacts is a complex task, since relationships, emotions and personality are, to a large extent, subjective and individual variables. Previous work has mainly focussed on anthropomorphic comparisons, assessing the naturalness of synthesised speech or of automatic talking-faces, as well as a general idea of user satisfaction (see, for instance, the evaluation session in Cassell *et al.* 2000). It is time now to develop an engineering-oriented evaluation framework, so as to ensure acceptability, comparability and utility of results. This section reports an initial proposal for the requirements for user interaction with Relationship Technologies. It is intended merely as a very preliminary contribution, essentially aimed at generating discussion and at laying a foundation for the involvement framework.

The concept of involvement refers to the strength and the quality of the relationship between a user — or a group of users — and a social artifact. Unlike usability, which has been mainly considered as a property of the interface, involvement is a relational property, generated by the encounter between two active agents: the user and the artifact. The difference is fundamental. Involvement is by definition a relative concept that strongly depends on the system's features, on the user's characteristics and on the level of familiarity between the two agents. Familiarity is a key concept and a clear component in the evaluation of social artifacts: It refers to the amount of personal knowledge about the partner which is available during the information exchange. Relationships unfold over time. Hence, familiarity is an evolving factor enhancing the control over the interaction: predicting the behaviour of a close friend is much easier than predicting the reaction of a stranger. Nevertheless, familiarity can have, as many realise, two opposite effects on the relationship: reinforcing it (attraction effect) or dissolving it (tedium effect).

Involvement is a multidimensional concept. Following the early results from our first study, we propose to divide it into the three dimensions referring to social, functional and aesthetic qualities. The relative weight of each dimension varies according to the task to be executed, the context of interaction and the nature, or personality, of the end-user. A simple representation of the involvement framework evaluating e-commerce social artifact is illustrated in Figure 7.4. The circles represent the dimensions and are labelled with the initial of the dimension to which they refer. The overlapping area stresses the blurred outlines between the dimensions. We assume that a number of mutual influences affect the social, aesthetic and functional qualities of any object. For this reason, user satisfaction is determined by the convergence of the perceived quality of each dimension.

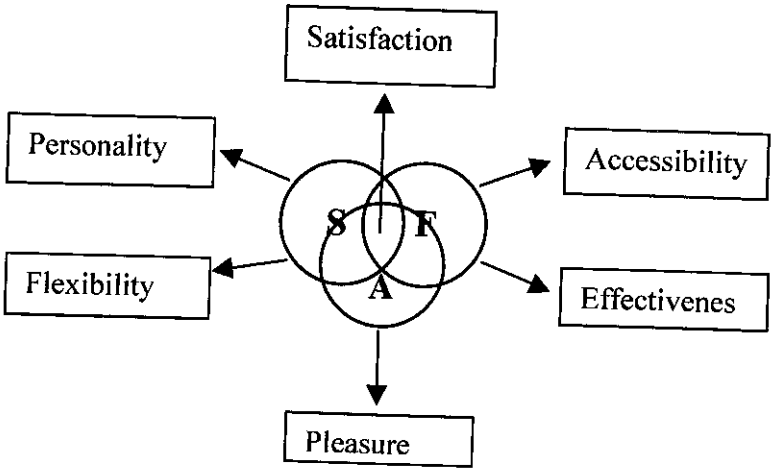

Figure 7.4 The involvement framework

Two basic functional qualities are considered relevant for successful social artifacts: Accessibility and Effectiveness. Accessibility refers to the effort required to interact with the social artifact, in terms of ease of communication and availability of the artifact via different technology channels mediating the same service(s). Natural Language (NL) appears to be a favourite interaction medium for social artifacts, since users can express their communicative intentions in a spontaneous way (Cassell *et al.* 2000). Moreover, NL interfaces have a tendency to elicit social representations (De Angeli *et al.* 1999).

Effectiveness reflects the key usability dimension. In this context, however, a number of indicators can be added to the traditional 'error rate'. The task of a social artifact is to provide a service whilst creating and maintaining a relationship with the user. Therefore, the 'return rate' and the 'amount of the bi-directional information flow' become fundamental variables to assess its overall success. The first refers to the number of times a customer interacts with the artifact. The second refers to the quantity and the quality of personal information disclosed by the user to the system, in addition to how many times the user acts on the information provided by the agent. The interaction time is a controversial variable. It refers to the temporal interval used to successfully exchange information with the system from the consumer's perspective. In the traditional usability model, time is often the major indicator of efficiency: the lower the time, the better the system. Social artifacts tend to invert this assumption: the higher the time, the better the system.

The major attribute underlying the aesthetic dimension is pleasure. It refers to the sense of enjoyment produced by all the aspects, and especially the physical aspect, of the interface. To maximise it, great care should be devoted to designing attractive and appealing interfaces. Moreover, the user should be encouraged to personalise the aspect of the social artifacts. Note that the aesthetic dimension overlaps with one of the basic social attributes, flexibility. Each form of successful communication is mediated by at least three factors: (a) the initial mental model of the conversational partner; (b) the evolution of this model over the course of communication; and (c) the ability to adapt to different partners. Hence, a social artifact must be capable of dynamically understanding the user and the context in which the interaction takes place, and in modifying its behaviour according to this knowledge.

A flexible system must be both adaptable (the user can directly personalise some

features according to her/his preferences) and adaptive (the system is capable of modifying its behaviour according to a user model). Traditionally, adaptivity in human-computer interaction has been focused on cognitive factors such as the knowledge, plans, interests, and preferences of the user. Here, the concept is extended to include extra-rational factors such as the user's attitudes, personality, or emotional state. In this sense, flexibility implies social intelligence (i.e., knowledge and understanding of shared social values and norms) and emotional sensitivity (i.e., the ability to recognise the emotional state of the conversational partner). It is evident that huge technical challenges must still be overcome before computers will be capable of easily or reliably understanding user's moods and desires. Nevertheless, affective computing (Picard 1997) is growing at a tremendous pace, which is likely to shape the future of computing (Ball and Breese 2000).

In addition to being flexible, a social artifact must exhibit a personality (a stable set of traits defining its overall character). Providing the system with a personality can increase the user's control over the interaction, and help to optimise the learnability of the system. To fulfil these objectives, the artifact must exhibit a consistent and stable personality. Unexpected and unpredictable swings between different attitudes can disorient the users and create a strong sense of discomfort. The system's personality must be predictable, both over time and across channels, for the same task (e.g. using a mobile 'phone' or an ATM to access a financial account balance, or personal information service).

In conclusion, we have argued that the current adherence to and reliance upon the accepted usability framework is challenged when we encounter social artifacts. We have created an exemplar 'interface agent' in the form of Granny and undertaken an initial study, which reveals great differences in people's perceptions of such an artifact, when compared with other actors/agents providing the same service. Finally, as a consequence of our study and the existing consensus on social artifacts, we have attempted to sketch the foundations of a framework based upon the construct of involvement, describing its components, which we believe offers a pragmatic means of evaluation. The next stages for our work in this area will be in the creation and evaluation of further social artifacts, manipulating their characteristics, and conducting empirical studies, as we develop the thinking behind the involvement framework. Our paper has taken us from a consideration of 'efficiency' as a key element of accepted usability definitions, to 'pleasure' as a component of involvement, a brief tour hosted in part by Granny.

7.7 ACKNOWLEDGEMENT

We are most grateful to NCR Corp. for their support, our colleagues in the Knowledge Lab for their encouragement and help with this project, to Sandra Ensby for her impressive work on the Granny visualisation, to Govert de Vries (now at in2sports.com) for his early input and enthusiasm, and to NCR Self-Service Advanced Solutions for their assistance in creating the ATM-Granny prototype.

7.8 REFERENCES

Baber, C., Johnson, G. I. and Cleaver, D., 1997. Factors affecting users' choice of words in speech-based interaction with public technology. In *International Journal of Speech Technology*, 2, 1, pp. 45-60.

Ball, G. and Breese, J., 2000. Emotion and personality in a conversational agent. In

Embodied conversational agents. Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. (Eds), The MIT Press, Cambridge, Mass.

Bernsen, N.O., Dybkjær, H. and Dybkjær L., 1998. *Designing interactive speech systems*. Springer Verlag, London.

Card, S., Moran, T. and Newell, A., 1983. *The psychology of human-computer interaction*. Erlbaum, Hillsdale, NJ.

Cassell, J., 2000. Embodied conversational interface agents. In *Communications of the ACM*, **43**, 4, April 2000, pp. 70-78.

Cassell, J., Sullivan J., Prevost, S. and Churchill, E., 2000. *Embodied conversational agents*. The MIT Press, Cambridge, Mass.

Costa, P.T. and McCrae, R.R., 1992. *NEO PI-R: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Coventry, L. and Johnson, G.I., 1999. More than meets the eye: Usability and iris verification at the ATM interface. In *Human-Computer Interaction: Interact 99*, Proceedings, **2**. Brewster, S., Cawsey, A. and Cockton, G. (Eds), IOS Press/IFIP, pp. 151-156.

De Angeli, A., 1997. *Valutare i sistemi flessibili: un approccio globale alla HCI*. Ph.D. Thesis, Psychology Dept., University of Trieste.

De Angeli, A., Gerbino, W., Nodari, E. and Petrelli, D., 1999. From tools to friends: Where is the borderline? In *Proceedings of the UM'99 Workshop on Attitude, Personality and Emotions in User-Adapted Interaction*. Banff, Canada, June 23, pp. 1-10.

De Angeli, A., Matera, M., Costabile, F., Garzotto, F. and Paolini, P., 2000. Validating the SUE Inspection Techniques. In *Proceedings of AVI2000*, June 23-27, Palermo, Italy.

Frøkjær, E., Hertzum, M. and Hornbæk, K., 2000. Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated. In *CHI 2000 Conference Proceedings*, ACM Press, NY, pp. 345-352.

ISO - International Standard Organisation, 1991. ISO 9241: *Ergonomics requirements for office work with visual display terminal (VDT)* - Parts 1-17.

Johnson, G.I., 1993. Spatial operational sequence diagrams in usability investigations. In *Contemporary Ergonomics 1993*. Lovesey, E.J. (Ed.), Taylor & Francis, London.

Johnson, G.I., 1995. Consumer reactions to voice at the ATM. In *Proceedings of the '95 Annual AT&T Human Factors & Behavioural Science Symposium*. AT&T Bell Labs, Holmdel, NJ, pp. 54-58.

Johnson, G.I., 1996. The usability checklist approach revisited. In *Usability evaluation in industry*. Jordan, P.W., Thomas, B., Weerdmeester, B.A. and McClelland, I.L. (Eds), Taylor & Francis, London.

Johnson, G.I., 1996b. Exploring novel banking user interfaces: Usability challenges in design and evaluation. Invited address, IEEE: *Interfaces – The leading edge*, C5 Group – Human Computer Interaction; Digest 96/126.

Johnson, G.I. and Briggs, P., 1994. Question-asking and verbal protocols – Qualitative methods of usability evaluation. In *Qualitative methods in organisation research: A practical guide*. Cassell, C. and Symon, G. (Eds), Sage Publications, London.

Johnson, G.I. and Westwater, M.G., 1996. Usability and self-service information technology: Cognitive engineering in product design and evaluation. In *AT&T Technical Journal*, Jan/Feb. AT&T, NJ, pp. 64-73.

Jordan, P.W., Thomas, B., Weerdmeester, B.A. and Clelland, I.L., 1996. *Usability evaluation in industry*. Jordan, P.W., Thomas, B., Weerdmeester, B.A. and McClelland, I.L. (Eds), Taylor & Francis, London.

Kelly, K., 1998. *New rules for the new economy*. Fourth Estate, London.

Lynch, P., De Vries, G., Johnson, G.I. and Waller, M., 1999. *Financial interfaces with personality*. Paper presented at the 3rd I^3 Conference, Affect in Interaction, October 20-

22nd, Sienna.

Lynch, P., Emmott, S.E. and Johnson, G.I., 1999. The NCR Knowledge Lab. In *Human-Computer Interaction: Interact 99, Proceedings*, **2**. Brewster, S., Cawsey, A. and Cockton, G. (Eds), IOS Press/IFIP, pp. 229-230.

Moon, Y. and Nass, C., 1996. How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. In *Communication Research*, **23**, 6, pp. 651-674.

Morkes, J. Kernal, H.K. and Nass, C., 1999. Effects of humour in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. In *Human-Computer Interaction*, **14**, pp. 395-435.

Nass C., Moon Y., Fogg, J., Reeves, B. and Dryer, D.C., 1995. Can computer personalities be human personalities? In *International Journal of Human-Computer Studies*, **43**, pp. 223-239.

Nass, C., Moon, Y. and Carney, P., 1999. Are respondents polite to computers? Social desirability and direct responses to computers. In *Journal of Applied Social Psychology*, **29**, 5, pp. 1093-1110.

Nass, C., Moon, Y. and Green, N., 1997. Are computers gender-neutral? Gender stereotypic responses to computers. In *Journal of Applied Social Psychology*, **27**, 10, pp. 864-876.

Nass, C. and Lee, K.M., 2000. Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In *CHI 2000 Conference Proceedings*, ACM Press, NY, pp. 329-336.

Nielsen, J., 1993. *Usability engineering*. Academic Press, San Diego, CA.

Norman, D.A., 1991. Cognitive artifacts. In *Designing Interaction: psychology at the Human-Computer Interface*. Carroll, J. M. (Ed.), University Press, Cambridge.

Picard, R., 1997. *Affective computing*. The MIT Press, Cambridge, Mass.

Reeves, B. and Rickenberg, R., 2000. The effects of animated character on anxiety, task performance, and evaluations of user interfaces. In *CHI 2000 Conference Proceedings*. ACM Press, NY, pp. 49-56.

Reeves, B. and Nass, C., 1996. *The media equation.* Cambridge University Press, New York.

Westwater, M.G. and Johnson, G.I., 1995. Comparing heuristic, user-centred and checklist-based evaluation approaches. In *Contemporary Ergonomics 1995*. Richardson, S. (Ed.), Taylor & Francis, London.

WordNet 1.6: available via http://www.cogsci.princeton.edu/