Persuasive Technology

AISB '08

AISB 2008 Proceedings Volume 3

GLoRiClass

6th Framework Programme

aisb ABERDEEN - 2008

UNIVERSITY OF ABERDEEN

AISB 2008 Convention
Communication, Interaction and Social Intelligence

1st-4th April 2008

University of Aberdeen

Volume 3:
Proceedings of the
AISB 2008 Symposium on Persuasive Technology

Published by
The Society for the Study of Artificial Intelligence and Simulation of Behaviour

http://www.aisb.org.uk/convention/aisb08/

ISBN 1 902956 62 1

Contents

The AISB'08 Convention: Communication, Interaction and Social Intelligence

As the field of Artificial Intelligence matures, AI systems begin to take their place in human society as our helpers. Thus it becomes essential for AI systems to have sophisticated social abilities, to communicate and interact. Some systems support us in our activities, while others take on tasks on our behalf. For those systems directly supporting human activities, advances in human-computer interaction become crucial. The bottleneck in such systems is often not the ability to find and process information; the bottleneck is often the inability to have natural (human) communication between computer and user. Clearly such AI research can benefit greatly from interaction with other disciplines such as linguistics and psychology. For those systems to which we delegate tasks: they become our electronic counterparts, or agents, and they need to communicate with the delegates of other humans (or organisations) to complete their tasks. Thus research on the social abilities of agents becomes central, and to this end multi-agent systems have had to borrow concepts from human societies. This interdisciplinary work borrows results from areas such as sociology and legal systems. An exciting recent development is the use of AI techniques to support and shed new light on interactions in human social networks, thus supporting effective collaboration in human societies. The research then has come full circle: techniques which were inspired by human abilities, with the original aim of enhancing AI, are now being applied to enhance those human abilities themselves. All of this underscores the importance of communication, interaction and social intelligence in current Artificial Intelligence and Cognitive Science research.

In addition to providing a home for state-of-the-art research in specialist areas, the convention also aimed to provide a fertile ground for new collaborations to be forged between complementary areas. Furthermore the 2008 Convention encouraged contributions that were not directly related to the theme, notable examples being the symposia on "Swarm Intelligence" and "Computing and Philosophy".

The invited speakers were chosen to fit with the major themes being represented in the symposia, and also to give a cross-disciplinary flavour to the event; thus speakers with Cognitive Science interests were chosen, rather than those with purely Computer Science interests. Prof. Jon Oberlander represented the themes of affective language, and multimodal communication; Prof. Rosaria Conte represented the themes of social interaction in agent systems, including behaviour regulation and emergence; Prof. Justine Cassell represented the themes of multimodal communication and embodied agents; Prof. Luciano Floridi represented the philosophical themes, in particular the impact on society. In addition there were many renowned international speakers invited to the individual symposia and workshops. Finally the public lecture was chosen to fit the broad theme of the convention – addressing the challenges of developing AI systems that could take their place in human society (Prof. Aaron Sloman) and the possible implications for humanity (Prof. Luciano Floridi).

The organisers would like to thank the University of Aberdeen for supporting the event. Special thanks are also due to the volunteers from Aberdeen University who did substantial additional local organising: Graeme Ritchie, Judith Masthoff, Joey Lam, and the student volunteers. Our sincerest thanks also go out to the symposium chairs and committees, without whose hard work and careful cooperation there could have been no Convention. Finally, and by no means least, we would like to thank the authors of the contributed papers – we sincerely hope they get value from the event.

Frank Guerin & Wamberto Vasconcelos

Persuasive Technology Symposium

INTRODUCTION

Can a web site persuade you to be politically active? Can a mobile phone motivate you to exercise? Does instant feedback on petrol use change how people drive? Do online rating systems inspire people to behave better online? This symposium focuses on how digital technology can motivate and influence people (or agents). It brings together researchers, designers, and developers interested in computers designed to change attitudes and behaviors in positive ways.

In a persuasive communication, a source tries to influence a receiver's attitudes or behaviours through the use of messages. Each of these three components (the source, the receiver, and the messages) affects the effectiveness of persuasion. In addition, the type of communication (the way the message is delivered) can impact a message's effectiveness. This symposium brings together researchers working on all these aspects of persuasion, from persuasive argumentation to persuasive user interfaces.

Persuasive technology has a great practical potential, for instance to improve health (encouraging a reduction in alcohol intake, smoking cessation, an increase in exercise, more healthy eating, and adherence to medical treatment) and to move towards sustainable living (encouraging a reduction in energy consumption, recycling, and use of public transport). There is a growing interest within the research community into persuasive technology, as shown by the emergence of the new Persuasive conference series (in Eindhoven, the Netherlands, 2006; Stanford, US, 2007; Oulu, Finland, 2008), as well as the successful series of workshops on Computational Models of Natural Argument (an area overlapping with persuasion).

TOPICS

Submissions were invited on all aspects of Persuasive Technology, including:

- Persuasive argumentation
    - Generating persuasive arguments (identifying discourse goals, choosing argument structure, content selection)
    - Ontologies for persuasion
    - Persuasive discourse processing: understanding what users say in terms of argumentation schemes
    - Computational models of argumentation
    - Rhetoric and affect: the role of emotions, personalities, etc. in models of argumentation.
    - Enhancing receiver involvement
- User modeling
    - Modeling receiver involvement
    - Modeling receiver position
    - Modeling personality and affective state for persuasion
    - Effect of cultural differences on persuasion
- Persuasive User Interfaces
    - Use of (multiple) Embodied Conversational Agents for persuasion
    - Communication settings (e.g. direct versus indirect communication)
    - Timing of persuasive messages/ when to interrupt the user
    - Effective presentation of arguments
    - Online dispute resolution
    - Mobile persuasion, persuasive images, persuasive video, persuasive games
- Peripheral routes of persuasion
    - Humor in persuasion
    - Enhancing source credibility
        - Building trust using natural language
        - Models of on-line trust/credibility
        - Effects of Source appearance, source similarity
- Alternative ways of persuasion
    - Using the influence of peers to persuade
    - Persuasion through incentives and punishment
- Evaluation methods for persuasive technology
- Ethics of persuasive technology
- Applications of persuasive technology, like in healthcare, education, e-commerce, politics

Accepted papers cover a wide range of these topics. In addition to presentations of accepted papers, there will be an invited talk by Ehud Reiter on "*Lessons learned from tailoring smoking cessation letters*", a panel on "*The Ethics of Persuasive Technology*" and discussions.

PROGRAM CHAIRS

Judith Masthoff, University of Aberdeen, UK
Chris Reed, University of Dundee, UK
Floriana Grasso, University of Liverpool, UK

PROGRAM COMMITTEE

Fiorella de Rosis, University of Bari, Italy
Nancy Green, University of North Carolina Greensboro, US
Helmut Horacek, University of the Saarland, Saarbrücken, Germany
Doug Walton, University of Winnipeg, Canada
Cees Midden, Eindhoven University of Technology, Netherlands
Peter de Vries, Twente University, Netherlands
Hien Nguyen, University of Aberdeen, UK
Susan Ferebee, University of Phoenix, US
Timothy Bickmore, Northeastern University, US
Jim Lyttle, Penn State Great Valley, US
Ingrid Zukerman, Monash University, Australia
Fabio Paglieri, ISTC-CNR, Rome, Italy
Marco Guerini, ITC-IRST, Povo-Trento, Italy
Guiseppe Carenini, University of British Colombia, Canada
Oliviero Stock, ITC-IRST, Italy
Katie Atkinson, University of Liverpool, UK

Using Digital Images to Enhance the Credibility of Information

Hien Nguyen[1] and **Judith Masthoff**[1]

Abstract. With research revealing the importance of trust as a deciding factor for users to visit a website, designing websites that users find credible becomes important. Adding onscreen characters in various forms (e.g. photographs of a person) to represent the source of information is a simple and popular way to to increase the credibility of a website. However, despite its popularity, empirical studies have reported mixed results. This paper presents three experiments that explored the causes of this contradiction. In our experiments, the credibility of the source (as resulting from a photograph of a person) was found to be topic dependent. It was also found to positively correlate with and account for some 30% of the variability in the credibility of a website. These findings perhaps can explain the inconsistency of many previous studies. Finally, our study suggested that adding a user-selected photograph increases the credibility of the website, while the opposite occurs when adding a photograph that is lowly credible with respect to the topic of the website's content.

1 INTRODUCTION

Social psychology has long emphasized the importance of the credibility (i.e. being trustworthy and knowledgeable) of the source (of information) in a persuasive communication. The credibility of the source is not a commodity that the source possesses, but it is the perception of the receiver about the source [1]. The more credible the receiver finds the source, the more credible the messages conveyed by the source become and vice versa [1]. It is acknowledged that one way used by the receiver to judge the source's credibility is through the source's appearance [2]. The source's appearance influences its credibility by affecting the receiver's perception of the underlying factors that form credibility. These factors commonly include *physical attractiveness*, *trustworthiness*, and *expertise* [1,2].

One way that web designers have tried to utilize this suggestion to enhance the credibility of (the information of) websites is by using onscreen characters to deliver the sites' information. This could be as simple as having a photograph of a person attached to the information, or as complex as having a fully animated character that is capable of exhibiting life-like behaviours such as speech, gestures, or expressions of emotions. However, despite its popularity, empirical studies have reported mixed results of this approach. While some studies highlighted the intended positive effect, others suggested otherwise. Two studies at Boston University showed that people were more willing to cooperate with a human-like character when that character had been made more attractive [3]. However, attractiveness alone was not sufficient to predict cooperation: subjects cooperated less with a more attractive, but dog-like character. Another study showed that adding a formal photograph of an author improved the trustworthiness, believability, perceived expertise, and competence of a web article (compared to an informal or no photograph) [4]. However, adding a photograph of a person did not enhance the perceived trustworthiness of a computer's recommendations [5]. Similarly, while Steinbrueck et al. [6] found that images of employees can increase users' trust in e-commerce sites, a later study by Riegelsberger [7] could not replicate this result. In his study, "neither the presence of a photo, nor the trustworthiness of the person depicted, had a significant effect" [7]. Other studies reported both positive and negative effects (e.g. [8]), or positive effects in some situations but not others (e.g. [9]). So what causes this inconsistency?

In this paper, we thoroughly explore this contradiction, focussing on one particular form of onscreen characters: *photographs of people*. Section 2 presents our hypotheses, while Section 3 discusses our experiments and their results with respect to the hypotheses postulated in Section 2. Finally, Section 4 summarises our findings and their implications.

2 RESEARCH QUESTIONS

In our earlier attempt to solve this contradiction [10], we argued that one reason for the inconsistency discussed in Section 1 is that the credibility of a person may well be topic dependent. When delivering information, a speaker might have high credibility in certain topics and low credibility in others. For instance, we can look at a user who is trying to be more physically active. In the user's situation, s/he might find a doctor more credible than a sport instructor to talk to about the benefits of exercise on health. In contrast, a sport instructor might have an edge over a doctor when talking about fitness programmes. Meanwhile, a close friend or someone who is similar to the user and has been in the user's situation might be the most credible person should the user need social support. This prediction led to our first hypothesis:

H1: *The (perceived) credibility of a person (as resulting from his/her appearance) is topic dependent.*

Indeed, H1 was confirmed by Experiment 1 in [10]. So perhaps, we did not see the expected results in many aforementioned studies because the selected photos were not highly credible with respect to the particular topic of the website that they were

[1] Dept. of Computing Science, Univ. of Aberdeen, AB24 3UE, UK. Email: {**n.nguyen,** `j.masthoff`}`@abdn.ac.uk`;

attached to (e.g. [5,7]), or the same photo was used across many websites promoting different products (e.g. [7,9]).

Therefore, we would expect that adding a photograph that is highly credible with respect to the topic of a website improves the credibility of the website. In our earlier study (see Experiment 2 in [10] for a complete discussion), we asked participants to judge the credibility of two web pages, one after the other. The first page informs the users about the health benefits of exercise, the second page provides guidance on planning personal fitness programmes. The web pages were presented to the participants with either a highly credible photo (HI, 19 participants), a lowly credible photo (LI, 19 participants) or without a photo (21 participants). Similar to other studies mentioned in Section 1, the credibility of the photographs was validated by a number of subjects prior to the experiment, and these subjects did not take part in the experiment.

	Source Credibility	Content Credibility
A web page about the health benefits of exercise		
HI (photo 7)	3.95	4.62 (0.84)
LI (photo 4)	3.02	3.90 (0.84)
No photo	-	4.30 (0.90)
A web page about planning personal fitness programmes		
HI (photo 12)	3.87	4.37 (0.74)
LI (photo 8)	3.25	4.07 (0.73)
No photo	-	4.32 (0.76)

Table 1. The results of our earlier study. Reported in [10].

The results of the experiment are summarised in Table 1. We found that attaching a photo with a high credibility increases the credibility of the first page that discusses the health benefits of exercise, while attaching a photo with a low credibility decreases its credibility (compared to the credibility of the page when no photo was present). The difference between the groups was statistically significant. The same trend was found for the second page that discusses how to plan personal fitness programmes. However, the magnitude of the increase and decrease of the credibility of the page were not statistically significant in this case. Some explanations could be offered to justify this result. First, the gap between the credibility of the two photos attached to the second page (0.62) was smaller than that of the two photos used in the first page (0.93), and this might have weakened the effect they had on the credibility of the page. Secondly, participants may not have taken the second part of the experiment as seriously, due to tiredness. Finally, it might well be that the participants who took part in the experiment did not judge the credibility of our highly credible photo as high, and our lowly credible counterpart as low as we had validated. Since credibility is a perception, it is unlikely that all viewers will perceive a particular photograph of a person equally. However, the right trend in our results suggests a positive correlation between the credibility of the photograph and the credibility of the content of the web page. This leads to our second hypothesis:

H2: *The perceived credibility of the person shown in the photograph with respect to a topic is positively related to the perceived credibility of the information of the website on that topic.*

If the credibility of a person is topic and receiver dependent as we pointed out, and if the credibility of a person on a topic positively correlates with the credibility of the information on that topic, then a user should find a message more credible if it is conveyed by a person that s/he personally finds highly credible with respect to that topic and vice versa. We also argue that, the user, when given the choice, will select the person whom s/he personally finds the most credible. This leads to our third hypothesis:

H3: *The perceived credibility of a message accompanied by a photograph chosen by the user for that topic is higher than that of the same message with no photograph, which in turn is higher than that of the same message with a lowly credible photograph.*

3 RESEARCH METHODOLOGY

Research Design

Our hypotheses are verified through a series of between-subjects experiments. Participants took part in the experiments voluntarily and no participant took part in more than one experiment.

Experiment 1: Source Credibility and Topic

Experimental Design

The aim of this experiment is two-fold: (1) to see whether the credibility of a person as judged from a photograph is indeed topic dependent, and (2) to validate the perceived age, gender, profession, and credibility for various potential photographs. While the former aims to test hypothesis 1, the later aims to find a subset of photographs with good inter-subject agreement on all these criteria to be used in future experiments. Details of the experiment were reported in [10], we will only briefly discuss this experiment here to enhance the continuity and understandability of this paper.

	5	4	3	2	1	
Attractiveness						
Attractive						Unattractive
Classy						Not classy
Beautiful/Handsome						Ugly
Elegant						Plain
Sexy						Not sexy
Trustworthiness						
Dependable						Undependable
Honest						Dishonest
Reliable						Unreliable
Sincere						Insincere
Trustworthy						Untrustworthy
Expertise						
Expert						Not an expert
Experienced						Inexperienced
Knowledgeable						Unknowledgeable
Qualified						Unqualified
Skilled						Unskilled

Table 2. Ohanian's credibility scale of a person.

In the experiment, the *credibility* of a person was defined as the average of his/her *attractiveness*, *trustworthiness*, and *expertness*, since these three factors are widely acknowledged to have the strongest influence on credibility [1,2]. A person's attractiveness, trustworthiness, and expertness were measured using 15 5-point Semantic Differential scale items developed by Ohanian [11] to measure a person's credibility. Each factor is calculated as the average of its five corresponding items (see Table 2 for exact wordings).

All participants were presented with 16 head and shoulder photographs of doctors and sport instructors / athletes. All photographs were taken from Microsoft Clipart (using search keywords like doctor and sport) and varied in age, and gender. The presentation order of the photographs was randomized for each participant to control for order effects. Examples of the photographs are shown in Figure 1. All the photographs can be found in [10].

Photo 7 Photo 4 Photo 12 Photo 8

Figure 1. Examples of photographs used in the experiments.

Participants were randomly assigned to one of four groups (to limit the time needed to perform the experiment and to avoid interaction effects between the questions). We asked each group to judge one or more characteristics of the person who appeared in each photograph, namely:

- Group A: *gender* (male or female), *age* (< 25, 25-30, 30-40, 40-45, or > 45), and *most likely profession* (doctor, sport instructor, or other),
- Group B: *attractiveness*
- Group C: *trustworthiness*
- Group D: *expertness* with respect to (1) the health benefits of exercise, and (2) fitness programmes.

Finally, all participants were presented with a web page showing all 16 photographs (the order of the photographs was also randomized for each participant). They could hover over each thumbnail to see the full size version. They were asked to choose whom they would like the most to learn from about each topic (i.e. health benefits of exercise and fitness programmes).

Participants

Fifty-one participants took part in the experiment (see Table 3 for the distribution of the participants' age and gender). Participants were staff and graduate students of the university, but came from all areas and professions.

	Gender		Age			
	F	M	18-20	21-24	25-29	30+
Number of subjects	32	19	5	17	9	20

Table 3. The distribution of participants' age and gender.

Results

Table 4 shows the average attractiveness, trustworthiness, and expertness of each photo. Hypothesis 1 stated that a person's credibility (as resulting from his/her appearance) is topic dependent. Indeed, all doctors were rated to have more expertise in the health benefits of exercise than in fitness programmes, while the opposite holds for all sport instructors, and two people who were identified as neither doctor nor sport instructor were rated to have the same level of expertise with respect to the two topics (also noted that their expertise was lower than that of both doctors and sport instructors). However, the difference between the levels of expertise in the two topics for sport instructors was much less significant than that of doctors. Perhaps, sport instructors are assumed to be interested in exercise not just for the sake of exercise, but also because they care about its benefits to health. Independent sample t-tests indicated that doctors are perceived as more expert with respect to the health benefits of exercise than sport instructors (average=3.92, stdev=0.21 vs. average=3.59, stdev=0.18, $p<0.05$), while sport instructors are perceived as more expert with respect to fitness programmes than doctors (average=3.86, stdev=0.22 vs. average=3.26, stdev=0.34, $p<0.05$). Thus, we concluded that H1 is supported in [10].

Photo	A	T	H Exp	#H	E Exp	#E
Photos which are perceived as doctors						
1	2.9 (0.7)	3.9 (0.8)	4.1 (0.6)	1	3.0 (0.9)	1
2	3.0 (0.9)	4.1 (0.7)	3.8 (1.1)	7	3.0 (0.8)	1
3	3.4 (0.6)	4.1 (0.8)	4.1 (0.6)	1	3.5 (0.5)	1
4	2.4 (0.6)	3.2 (0.9)	3.5 (1.3)	1	2.5 (1.1)	0
5	2.7 (0.6)	4.6 (0.5)	4.0 (0.8)	1	3.4 (0.7)	1
6	2.7 (0.8)	3.7 (0.7)	3.9 (0.7)	2	3.5 (0.6)	3
7	3.3 (0.7)	4.5 (0.5)	4.1 (0.8)	8	3.5 (0.6)	2
8	2.5 (0.7)	3.8 (1.0)	4.0 (0.9)	1	3.4 (0.6)	0
11	3.0 (0.8)	4.1 (0.7)	3.9 (0.6)	1	3.5 (0.7)	0
Photos which are perceived as sport instructors / athletes						
9	3.8 (0.8)	3.2 (1.1)	3.7 (0.6)	5	3.9 (0.5)	14
10	2.8 (0.9)	3.0 (0.9)	3.6 (1.0)	6	4.0 (0.7)	9
12	3.6 (0.6)	3.9 (0.9)	3.8 (0.7)	5	4.1 (0.5)	12
13	2.8 (0.9)	4.3 (0.5)	3.3 (0.9)	1	3.7 (0.9)	2
15	2.6 (0.9)	4.3(0.6)	3.5 (0.8)	7	3.6 (0.8)	3
Photos which are perceived as other						
14	2.4 (0.5)	3.3 (1.2)	3.0 (0.9)	3	3.1 (1.1)	2
16	2.5 (0.9)	3.8 (0.9)	3.2 (1.2)	1	3.2 (1.2)	0

* **A** = Attractiveness; **T** = Trustworthiness.

H Exp, **E Exp** = Expertness on the topic of health benefits of exercise, fitness programmes respectively.

#H, **#E**: number of subjects who picked this photo as their favourite to learn from about the health benefits of exercise and fitness programmes respectively.

Table 4. Average score and standard deviation of each photo.

Experiment 2

Experimental Design

As we previously discussed in section 2, there ought to be a positive correlation between the credibility of a web page and the credibility (with respect to the topic of the web page) of the photo attached to it (H2). This experiment aims to verify whether this correlation does in fact exist.

This experiment is an improvement of our earlier study that we discussed in section 2 (also see [10, Section 3] for a complete discussion). All participants were shown the same web page on how to plan a personal fitness programme that was used in the previous study. We composed the content of the web page based on the information available on www.mayoclinic.com[2]. Participants were asked to read the given web page carefully. After reading the page, they were asked to judge its credibility. The credibility of the page was measured using 15 7-point Likert scale items. The items were developed and validated by Hong [12] for assessing the credibility of health-related websites. They assess five commonly recognised dimensions of the *credibility* of information: *expertise*, *goodwill*, *trustworthiness*, *depth*, and *fairness* (see Table 5 for exact wordings). Credibility is calculated as the average of the five dimensions. Each dimension is the average of its respective items. The items were ordered such that no two items from the same dimension appeared sequentially.

Fairness
This page provides information that is neutral
This page provides information that is **not** balanced
This page is biased in the information it provides
This page is slanted in the information it provides
This page is even-handed in presenting information
Depth
This page does **not** provide in-depth information
This page is **not** comprehensive
This page offers everything you need to know on the topic
Goodwill
This page has my interests at heart
This page is uncaring about its visitors
This page is **not** concerned about its visitors
Trust/Expertise
This page appears to have experts on the topic discussed
This page is ethical
This page appears to be a leader in its area of specialty
This page is **not** trustworthy

Table 5. Hong's credibility scale for a health-related web page.

All participants also indicated the extent to which they already knew the information presented to them, their knowledge on the topic, and their reliance on the Internet for seeking health-related information. The exact wordings and results are shown in Table 7. Note that the first three questions were asked before, and the last two after the participants read the message.

Participants were randomly assigned to one of three groups. The message for the first (the Highly Credible Image (HI)) and second (the Lowly Credible Image (LI)) group prominently showed a highly credible photo (photo 12) and a lowly credible photo (photo 8) respectively. The photos were the same as the ones used in our earlier study (see Table 6). The participants were told that the source of the content of the page was the person who appeared in the photograph, and asked to judge the credibility of the person after they judged the credibility of the page. The credibility of the person was measured in the same way used in Experiment 1. The message for the third group (the No Image (NI)) did not have any photo attached. Participants in this group were also asked to judge the credibility of the source of the message after they judged the credibility of the message. For the NI group, we used the same measures used in Experiment 1 except that the questions about attractiveness were taken out since they were not applicable.

	 Photo 12	 Photo 8
Attractiveness	3.63	2.52
Trustworthiness	3.85	3.83
Expertness in fitness programs	4.13	3.38
Overall Credibility for fitness programmes	**3.87**	**3.25**

Table 6. The photographs used in the experiment.

Participants

Thirty-one participants took part in the experiment (see Table 7 for the participants' demographics).

	HI		LI		NI	
Number of participants	11		11		9	
Gender	M	F	M	F	M	F
	5	6	7	4	5	4
Average age	30.00 (15.88)		34.27 (8.43)		30.11 (11.86)	
I am currently doing some form of exercise	Y	N	Y	N	Y	N
	6	5	8	3	8	1
I am more educated about my health than most people*	4.91 (0.83)		4.91 (1.30)		4.44 (1.74)	
I have full knowledge of the benefits (consequences) of regular exercise (the lack of it)*.	5.55 (1.04)		5.45 (1.44)		5.56 (1.88)	
When I need information about benefits of exercise and fitness programmes, I would go to the Internet*.	4.73 (1.68)		4.91 (1.22)		4.33 (2.06)	
I already knew all the information presented**	5.18 (1.08)		4.91 (1.51)		4.89 (2.32)	

* 1 = strongly disagree 7 = strongly agree
** 1 = nothing 7 = everything

Table 7. The participants' demographics.

Participants were recruited online from an experiment listing website maintained by the University of Zurich

[2] The content of the web page can be found at www.csd.abdn.ac.uk/~hnguyen/aisb08/fitness.txt

(http://genpsylab-wexlist.unizh.ch/), and from a mailing list that staff and students of the university subscribed to. Participants came from all areas and professions.

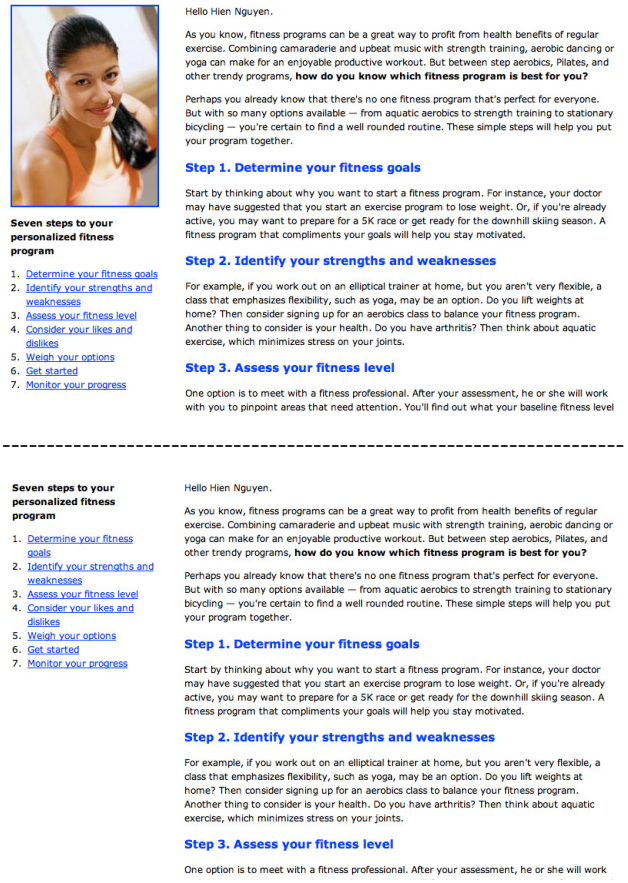

Figure 2. The website with and without photo.

Results

Hypothesis 2 stated that the perceived credibility of the person shown in the photograph with respect to a topic positively correlates with the perceived credibility of the information of the website on that topic. Table 8 shows the Pearson correlation test results.

Overall, the credibility of the source is positively related to the credibility of the message when a photograph of the source is present ($r=0.557$, $R^2=0.310$, $p<0.01$). This result also pointed out that the credibility of the source accounts for some 31% of the variability in the credibility of the message, hence it would not be the only factor influencing the users' perception of the message credibility, especially when considering how simple our site was (see Figure 2).

This might give a good explanation for the non-significant correlation between the source and the message credibility found in the NI and LI group. We noted that the majority of the participants in LI and NI groups are currently exercising, which was not the case for the HI group. This may have weakened the impact of the credibility of the source on the credibility of the message, leading to the non-significant correlation. A follow-up experiment will be needed to investigate these issues further.

		Source			
		Attractiveness	Trust	Expertise	**Overall credibility**
Web page credibility	HI	0.143	0.853*	0.733*	**0.695***
	LI	0.098	0.395	0.296	**0.313**
	With photo (HI+LI)	0.125	0.710*	0.544*	**0.557***
	NI	n/a	0.554	0.233	**0.446**

* Correlation is significant at the 0.01 level (1-tailed)

Table 8. Pearson's correlations between the credibility of the source and the credibility of the web page.

Experiment 3

Experimental Design

This experiment is designed to verify our third hypothesis, which stated that adding a user-selected photograph increases, while adding a lowly credible photograph decreases the perceived credibility of the information.

All participants were asked to select the reasons why they find regular exercise difficult from a set of ten common reasons (e.g. "I don't have enough time to exercise", "I'm too tired to exercise after working all day"). For each reason, the participants can rate its importance on a 5-point Likert scale (from "not at all" to "strongly agree"). Participants were then presented with either a standard or personalised message containing suggestions to their problems (e.g. "Claim the back row of the parking lot as your own. Or park a few blocks away and walk quickly to your destination"). They were asked to read the given information carefully. They were then asked to judge the credibility of the message and the source. These were set up in the same way used in Experiment 2.

All participants also indicated the extent to which they already knew the information presented to them, their knowledge on the topic, and their reliance on the Internet for seeking health-related information. The questions were the same as those used in Experiment 2.

Participants were randomly assigned to one of four groups:

(1) Personalised Image and Personalised Content (PI+PC).
(2) Lowly Credible Image and Personalised Content (LI+PC).
(3) No Image and Personalised Content (NI+PC).
(4) No Image and No Personalised Content (NI+NPC).

Participants in the first group (PI+PC) were asked to choose someone whom they would like to speak to by clicking on one of the sixteen photos used in Experiment 1, prior to selecting reasons for finding regular exercise difficult. They were then presented with a personalised message. The chosen image was prominently shown alongside the message. The message was personalised to only contain suggestions to problems that they admitted having, and the suggestions were arranged in order of their importance (i.e. suggestions to problems rated 5 were put first). Suggestions to problems that were rated "not at all" were not included. The second group (LI+PC) received the same personalised message as the first group did but they could not choose which person they would like to talk to. They were given

the lowly credible image used in Experiment 2 (photo 8). The third group (NI+PC) also received the same personalised message as the first group did but the message was shown without any image. The fourth group (NI+NPC) received a non-personalised message containing suggestions to all ten problems in a random order regardless of what they had selected. The message was shown with no image attached.

Participants

Forty-four participants took part in the experiment. The participants' demographics are shown in Table 9 below. Participants were recruited online from an experiment listing website maintained by the University of Zurich (http://genpsylab-wexlist.unizh.ch/), and from a mailing list that staff and students of the university subscribed to. Participants came from all areas and professions. To guarantee that the participants replicate the actual potential users, we asked the participants whether they "currently want to exercise more regularly, but find it difficult" before the start of the experiment. Only participants who answered yes to this question were selected to complete the experiment.

	PI+PC		LI+PC		NI+PC		NI+NPC	
Number of participants	12		10		11		11	
Gender	M	F	M	F	M	F	M	F
	3	9	3	7	2	9	3	8
Average age	27.92 (13.70)		29.30 (7.63)		38.09 (12.05)		32.00 (11.95)	
Exercising[1]	Y	N	Y	N	Y	N	Y	N
	9	3	4	6	5	6	6	5
Education[2]	5.08 (1.16)		4.80 (1.62)		5.45 (1.21)		4.18 (0.98)	
Knowledge[3]	5.50 (0.90)		5.60 (1.26)		5.82 (1.54)		5.27 (1.62)	
Info known[4]	4.83 (1.19)		4.70 (1.34)		4.36 (1.96)		3.82 (1.94)	
No of items read	6.42 (2.57)		7.60 (2.91)		6.36 (2.77)		10 (0.00)	

[1] I am currently doing some form of exercise

[2] I am more educated about my health then most people (1 = strongly disagree 7 = strongly agree)

[3] I have full knowledge of the benefits (consequences) of regular exercise (or the lack of it) (1 = strongly disagree 7 = strongly agree)

[4] I already knew all the information presented (1 = nothing 7 = everything)

Table 9. The participants' demographics.

Results

The data collected in this experiment was also used to verify hypothesis 2, which postulated a positive correlation between the credibility of the person on a topic and the credibility of the message on that topic. Table 10 shows the Pearson correlation test results. The results are stronger than those of Experiment 2. The credibility of the source positively correlated with the credibility of the message when a photo is shown ($r=0.619$, $R^2=0.383$, $p<0.01$) as well as when no photo is present ($r=0.732$, $R^2=0.536$, $p<0.01$). Thus, H2 is supported. Interestingly, when looking at each individual group, the correlation between the source's and the message's credibility was not significant for the PI+PC group, and the participants' demographics (Table 9) showed that the majority of people in this group are currently exercising. This is highly similar to our findings in Experiment 2.

		Source			
		Attract-iveness	Trust	Expert	**Overall Credibility**
Web page credibility	PI+PC	0.349	0.479	0.326	0.462
	LI+PC	0.426	0.218	0.541	0.641*
	With Photo	**0.424***	**0.471***	**0.475***	**0.619****
	NI+PC	n/a	0.542*	0.785**	0.716**
	NI+NPC	n/a	0.490	0.749**	0.758**
	No Photo	**n/a**	**0.513****	**0.763****	**0.732****

* Correlation is significant at the 0.05 level (1-tailed)
** Correlation is significant at the 0.01 level (1-tailed)

Table 10. Pearson's correlation between the credibility of the source and the message.

Hypothesis 3 stated that the presence of a personalised photo increases and the presence of a lowly credible photo decreases the credibility of the message. Similarly to previous experiments, we averaged the results on each dimension to get an overall score of credibility. The results of the experiment, shown in Table 11 below, reflect our prediction perfectly. The perceived credibility of the message when accompanied by a personalised photo was the highest, followed by the non-personalised message and the personalised message with no photo, and at last the personalised message with a lowly credible photo. There were, however, two unexpected results.

	F	D	G	T/E	**Cred**
PI + PC	5.03 (0.89)	3.97 (0.12)	5.69 (0.82)	4.77 (0.59)	**4.87 (0.48)**
LI + PC	4.48 (0.63)	3.23 (1.36)	5.07 (0.78)	3.98 (1.06)	**4.19 (0.49)**
NI + PC	4.84 (0.97)	3.52 (1.44)	5.27 (1.03)	4.50 (1.18)	**4.53 (0.85)**
NI + NPC	4.25 (1.10)	4.03 (1.24)	5.24 (1.59)	4.77 (0.93)	**4.58 (0.83)**

* **F** = Fairness, **D** = Depth, **G** = Goodwill, **T/E** = Trust/Expertise, **Cred** = Credibility

Table 11. Average score and standard deviation of the perception of message credibility of each group.

First, only the difference between the PI+PC and the LI+PC group was statistically significant (independent t-test, $p<0.01$). So, the credibility of the website with an image selected by the user was not significantly higher than that of the website without any image at all. There are two plausible explanations for this unexpected result. Firstly, participants may not have regarded the person in the chosen picture as the source of the message. We assumed that asking participants to choose someone whom they would like to talk to prior to presenting them a message with the photo of the selected person would give the illusion that the picture is the source of the message. However, choosing a picture and then finding it attached to the message is perhaps not the same as naturally finding a credible image already attached

to the message. The fact that subjects had to select from a large set of images (16) may well have made it less likely that they would assume there would be so many different contents.

A second possible explanation for the unexpected result is that participants may not have selected the most credible person for them. Indeed, while the average credibility of the person in the user-selected photo was significantly higher than that of the lowly credible photo (average=3.56, stdev=0.38 vs. average=3.29, stdev=0.41, p=0.05), it was surprisingly not higher than that of the source when no photo was included (average=3.56, stdev=0.38 vs. average=3.76, stdev=0.70 (personalised message), and average=3.72, stdev=0.66 (standard message)). Only three out of twelve participants chose photo 12 or 10, which were the most favourite people to learn from about fitness programmes in Experiment 1 (see Table 4). Other chosen photos were: photo 6 (3 participants), photo 16 (1), photo 5 (3), photo 3 (1), and photo 4 (1). In addition, an average credibility of 3.56 for the person in the user-selected photo was considerably lower than what we would have expected considering the most credible person (photo 12) in Experiment 1 had a credibility of 3.87. Perhaps participants did not take the task of selecting a picture very seriously, knowing this was only an experiment, and suspecting that the person chosen would not influence the message.

These issues could be overcome by a more effective and unobtrusive way for identifying a user's preference for the appearance of the source. We will also consider reducing the number of images the user selects from, and more explicitly mentioning that the text will change depending on the person selected.

A second unexpected result is that our personalised content did not result in any improvement of credibility over the standard content. With hindsight, this is most likely due to the way personalisation is done: the standard content was longer (showing all ten information items) than the personalised content which only showed a subset (on average about three items less). Longer content is likely to result in a better rating for Depth (and there is a trend in that direction when comparing the NI+PC group with the NI+NPC group). Actually, studies in the area of recommender systems have shown that users appreciate longer reviews more [e.g. 18]. A more controlled follow-up study is needed to see if personalising the selection of and order in which the information items are presented has an effect when content length remains the same.

4 DISCUSSION

In this paper, we systematically investigated the effect of photographs of people on the perceived credibility of (the information) of a website.

Limitations

To avoid any overgeneralization of our conclusions, we will first discuss the limitations of the study. Firstly, the majority of our participants are female, and over 30 years old. Hence, any findings in this paper should only be generalized to other demographics with care. Secondly, our findings are restricted to the subject of supporting people to exercise more regularly and planning a fitness programme.

Findings and Implications

(1) The credibility of a photo of a person is topic dependent.

Firstly, Experiment 1 showed that *the perceived credibility of a person as resulting from his/her appearance is topic dependent* (H1). The same person is likely to have different credibility when conveying information about different topics. This finding might explain the inconsistent results of a number of studies (e.g. [7, 9]), in which a photograph was assumed to have the same credibility when being used on different sites specialising in different products (digital cameras, computer hardware, and flower services in [7], and digital cameras, vitamins, and tour packages in [9]).

(2) The more credible the person in the photo, the more credible the website. Adding a lowly credible photo can have a negative effect.

Secondly, we found a positive correlation between the credibility of the person in a photograph and the credibility of the website, and it accounts for some 30% of the variability in the website's credibility. So while the credibility of the person in the photograph is not the only factor that can influence the perception of the website's credibility, it clearly has an undeniable role. So while *adding the right image can positively affect the perception of trust*, *adding a lowly credible image can have a negative effect on the perceived credibility of the website*. These findings perhaps can further explain the mixed results found in other experiments where the effects of other variables might have not been accounted for, especially when the websites were more complicated.

(3) Choosing the right photo is crucial yet challenging

It is trivial to see from our findings the importance and difficulty of choosing the right photo. But how can we do this?

Further qualitative analysis of Experiment 1 revealed that the participants apply different criteria in choosing someone who they would like the most to talk to about a certain topic. The most mentioned ones were "fit (but not overly fit)" (16/51 people), "expert" (16/51), friendly (6/51). These criteria are not only different between participants but also between the topic of the health benefits of exercise and planning personal fitness programmes (see [10, Section 2] for a more complete analysis). In fact, only 7 out 51 participants chose the same favourite person for the two topics. In a series of studies by Baylor [13] where participants selected a human-like agent to act as their teacher, participants had the tendency to select an agent that was highly similar to themselves in terms of gender and ethnicity. These user-selected agents were perceived to be more credible and motivating than a standard agent or no agent. An interview of thirty Internet shoppers by Keeling [14] also revealed that what is considered the appropriate appearance for an agent attached to an e-commerce website not only depends on the users but also the product sold on the site. This suggests that *finding a generic image that can be liked and perceived as highly credible by the majority of users would be considerably challenging, especially when there are great varieties in the user population*.

There is a clear need for an effective and unobtrusive method for identifying a user's preferences for the appearance of onscreen characters. Our initial attempt of simply asking the

users to select someone whom they would like to speak to about a certain topic did not work as well as expected. On a positive note, adding a user-selected photograph yielded a significant improvement on both the perceived credibility of the source and that of the information over adding a lowly credible photograph with respect to the topic of the website. However, although adding a user-selected photograph enhanced the credibility of the website when no photo was present, the difference was not significant. Was it because the participants did not choose the most credible photo in the experimental condition? Was it because selecting a photo and then finding it attached to a message does not lead to the perception that the chosen photo is the source of the message? The responses from the participants in Keeling's study [14] showed that the participants indeed "varied in the extent to which they regarded the ECA as presenting the information". These issues will be addressed in a future study.

Conclusions

In summary, our findings clearly influence the process of using photographs (and onscreen characters in a broader sense) to improve the credibility of websites. While adding such characters can have a positive impact, choosing the right appearance is crucial, as the wrong appearance can lead to the opposite of the intended effect.

Future Work

Firstly, the vast majority of studies regarding the relationship between displaying photographs of people on websites and the sites' credibility, including our own, do not investigate how this relationship develops over time. Adding a photograph might be able to create a good first impression, but will it last? How will it increase or decrease over multiple interactions? We plan to explore this question further.

Secondly, we plan to compare how different forms of a source's visual appearance (e.g. animated characters, photographs of people, videos of people, voice, etc.) affect the credibility of a website.

Thirdly, we plan to further investigate the tailoring of images to users. As mentioned above, our first attempt of letting users select their favourite person to talk to was not very successful. We will investigate whether we can obtain better results by allowing users to select a person from a smaller set and more explicitly telling users that the content will change depending on the person they select. We will also investigate whether automatically selecting an image based on user stereotypes has a positive effect.

Finally, the use of onscreen characters is not restricted to having one character interacting and delivering the information to the users in a direct communication setting. For instance, multiple characters can be used to simulate the positive effect of social norms, and indirect communication on the perception of the credibility of the information [1,2]. Such impact is worth investigating, yet there have been few studies addressing it [15-17].

ACKNOWLEDGEMENTS

Judith Masthoff's research is partly supported by Nuffield Foundation Grant No. NAL/00258/G. We would like to thank the anonymous reviewers for their improvement suggestions.

REFERENCES

[1] J. B. Stiff. *Persuasive Communication*. The Guildford Press, New York, (1994).

[2] E. Aronson. *The Social Animal*. Worth, (2004).

[3] S. Parise, S. Kiesler, L. Sproull, and K. Waters. Cooperating with Life-like Interface Agents. *Computers in Human Behavior*, Vol 15 No 2, 123-142 (1999).

[4] B. J. Fogg, J. Marshall, T. Kameda, J. Solomon, A. Rangnekar, J. Boyd, and B. Brown. Web Credibility Research: A Method for Online Experiments and Early Study Results. *CHI 2001*, 295-296 (2001).

[5] P. De Vries. Social Presence As A Conduit to The Social Dimensions of Online Trust. *Persuasive Technology Conference*. Eindhoven, The Netherlands, 55-59 (2006).

[6] U. Steinbrueck, H. Schaumburg, S. Duda, and T. Krueger. A Picture Says More Than A Thousand Words – Photographs As Trust Builders In E-Commerce Websites. *In Proceedings of CHI 2002: Extended Abstracts*. Minneapolis, MN, US, 748-749 (2002).

[7] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy. Shiny Happy People Building Trust? Photos on E-commerce Websites and Consumer Trust. *Proceedings of CHI 2003*, 121-128 (2003).

[8] J. Riegelsberger, and M. A. Sasse. Face It: Photographs Don't Make Websites Trustworthy. *In Proceedings of CHI2002: Extended Abstract*, Minneapolis, MN, US, 742-743 (2002).

[9] P. De Vries, and A. Th. H. Pruyn. Source Salience and The Persuasiveness of Peer Recommendations: The Mediating Role of Social Trust. *In Proceedings of the 2nd Persuasive Technology Conference*. California, CA, US, Springer Verlag, 164-175 (2007).

[10] H. Nguyen, and J. Masthoff. Is It Me or Is It What I Say? Source Image and Persuasion. *In Proceedings of the 2nd Persuasive Technology Conference*. California, CA, US, Springer Verlag, 231-242. (2007).

[11] R. Ohanian. Construction and Validation of A Scale To Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, Vol 19 No 3, 39-52 (1990).

[12] T. Hong. Contributing Factors To The Use of Health-related Websites. *Journal of Health Communication*, Vol 11, 149-165 (2006).

[13] A. L. Baylor, R. B. Rosenberg-Kima, and E. A. Plant. Interface Agents as Social Models: The Impact of Appearance on Females' Attitude Toward Engineering. *CHI 2006*, Montreal, Canada (2006).

[14] K. Keeling, S. Beatty, P. McGoldrick, L. Macaulay. Face Value? Customer Views of Appropriate Formats for Embodied Conversational Agents (ECAs) in Online Retailing. *Proceedings of the 37th Hawaii International Conference on System Sciences*, (2004).

[15] S. D. Craig, B. Gholson, M. H. Garzon, X. Hu, W. Marks, P. Wiemer-Hastings, Z. Lu. Autor Tutor and Otto Tudor. *AIED-Workshop on Animated and Personified Pedagogical Agents*, Le Mans, France, 25-30 (1999).

[16] E. Andre, T. Rist, S. van Mulken, M. Klesen, S. Baldes. The Automated Design of Believable Dialogues for Animated Presentation Teams. *In J. Cassell, S. Prevost, J. Sullivan, and E. Churchill: Embodied Conversational Agents*, The MIT Press, 220-255 (2000).

[17] H. Nguyen, J. Masthoff, P. Edwards. Persuasive Effects of Embodied Conversational Agent Teams. *Proceedings of 12th International Conference on Human-Computer Interaction*, Beijing, China, Springer-Verlag, Berlin, 176-185 (2007).

[18] Tintarev, N. and Masthoff, J. (2007). Effective explanations of recommendations: User-centered design. Proceedings of the ACM Conference on Recommender Systems, Minneapolis, US, 153-156.

The persuasive effects of positive and negative social feedback from an embodied agent on energy conservation behavior

Cees Midden[1] **& Jaap Ham**[1]

[12]

Abstract. In this paper we explore the persuasive effects of social feedback, as provided by an embodied agent, on behavioral change. In a lab setting an experiment was conducted in which participants had the opportunity to conserve energy while carrying out washing tasks with a simulated washing machine. The experiment tested the effect of positive and negative social feedback and compared these effects to more widely used factual feedback. The robotic source, a so-called iCat, was manipulated with regard to its perceived agency. The results showed an effect of social feedback compared to factual feedback. In addition an effect of feedback valence was found, demonstrating more conservation actions following negative social feedback. The predicted perceived agency effect could not be demonstrated.

1 INTRODUCTION

The exhaustion of natural resources and the threats of growing CO2-emissions and climate change effects have urged nations worldwide to seek for substantial reductions in energy consumption. Although technological solutions like more efficient systems and devices and the development of renewable energy sources are of great importance, consumer behavior plays a crucial role in bringing down the level of energy consumption.

Influencing consumer behavior to promote energy conservation has become an important target of national and international policy efforts. Thereby, the question which instruments should be applied to promote energy conservation behavior has become highly relevant.

Recent reviews (e.g. Abrahamse, Steg, Vlek, & Rothengatter, 2005; Midden, Kaiser, & McCalley, 2007) have evaluated the effects of interventions to promote energy efficient behavior. In general, mass media public campaigns seem to lack precision in targeting and message concreteness to achieve behavioral change. By contrast, raising people's awareness of energy consumption by providing tailored feedback about their energy consumption (for example in kWh) can promote the achievement of behavioral change. The results are mixed though. Weak linkages between specific actions and energy outcomes caused by low feedback frequencies (e.g. once month) and insufficient specificity of the feedback (e.g. household in general vs. specific person or specific devices) are underlying these findings. Recently, technological solutions have created new opportunities to improve feedback efficacy by embedding it in user-system interactions. That is, energy use is in essence always the outcome of an interaction between a user and some energy using device. Intervening in these specific interactions might improve the quality of feedback substantially. Some evidence supports this claim. McCalley and Midden (2002, 2006) demonstrated in several studies that interactive forms of feedback can be effective to enhance energy efficient use of devices like a washing machine. By adding an energy meter to the user interface of a washing machine they achieved 18% of energy conservation both in lab and field studies. Basically, their approach entailed giving factual feedback in terms of kWh consumed as a function of programming choices made by the user, like water temperature, spinning speed or the duration of the washing cycle. In the lab study the subjects made choices on a simulated washing machine panel (see figure 1). In the feedback condition a vertical bar was added that indicated the relative level of energy consumed. The scale referred to kWh consumed but the exact amounts were not indicated. In the field real machines were installed in family homes either with or without the feedback bar on the panel. This study confirmed the lab findings.

In the present study we follow up on this work by aiming to increase the persuasiveness of the same system through the introduction of *social* feedback. We examine whether social feedback can add to the promotion of pro-environmental behaviors such as energy conservation in the home.

Social reinforcement has been applied widely in many domains such as child education, therapeutic programs, health behavior and social interaction as a mechanism for behavioral change (see e.g., Bandura & MacDonald, 1963). Social praise and compliments operate as positive incentives. Negative social incentives, like signs of disapproval, have been applied less widely, but nonetheless have been demonstrated to be effective as well (e.g. Wright, 1968).

Assuming the effectiveness of social reinforcements as delivered by human actors, can we expect social reinforcement to work as well when submitted by an intelligent system? The research on the media equation (Reeves & Nass, 1996) suggests that people show similar social behavior in interaction with computer systems as with humans. Social praise from a computer has been reported to enhance the attractiveness of the computer, and made people feel better and more positive about the interaction with the

[1] Eindhoven University of Technology, Netherlands; contact address: c.j.h.midden@tue.nl/j.r.c.ham@tue.nl

[2] We thank Susanne Tak and Teddy McCalley for their support with programming the feedback and preparing the user interface, Frans Jansen for his help with running the experiment, and Jan-Roelof de Pijper and Martin Boschman for their technical support.

computer (Fogg & Nass, 1997). In spite of the relevance of these findings, no direct evidence is available for the effects of social reinforcements on *behavioral change*.

In the present research we want to explore the effectiveness of social feedback as provided by a smart robotic agent and compare this with the more widely used factual forms of feedback, in particular interactive feedback, which can be considered as the most successful type (Midden, Kaiser & McCalley, 2007).

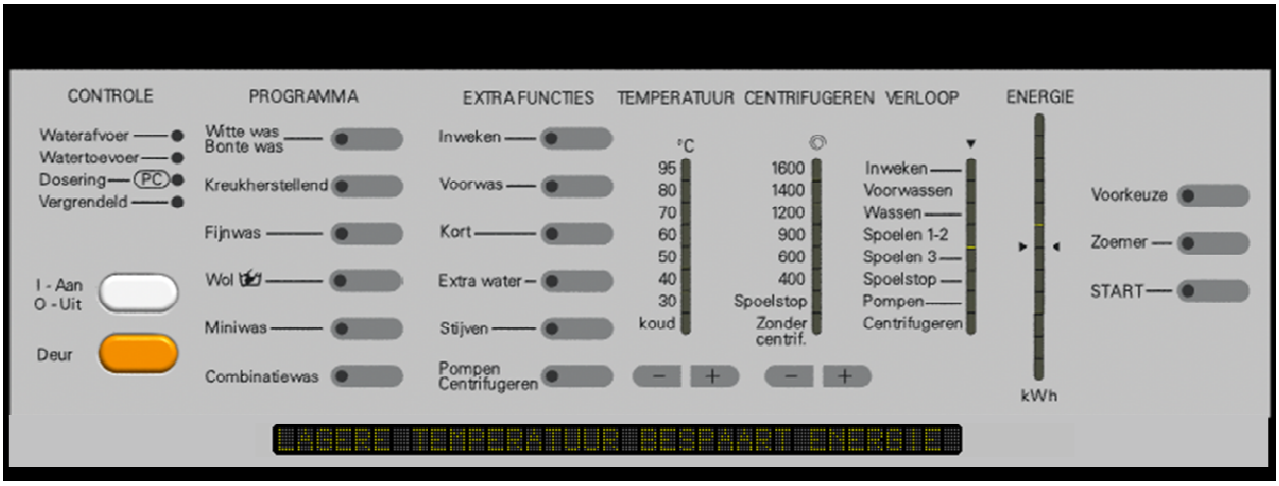

Figure 1: Washing machine interface with energy bar

The media equation studies have shown that people are not very sensitive to social differences between computers and humans. At least people seem to react similarly when confronted with computer or human agents. Social cues seem to automatically activate social reactions. Considering the finding by Fogg and Nass, we wanted to explore whether these conclusions would also hold for *behavior change* interventions through social feedback. It seems plausible that an intelligent system that is perceived as social actor should also be able to influence the behavior of a person that receives social feedback from the system.

Surprisingly, social feedback by smart computer agents has mainly focused on positive social feedback only. In spite of the motivational effects that have been observed of positive feedback, there are reasons to believe that the value of negative feedback may have been underrated. Research on the effects of positive and negative messages has shown that the impact of negative information can be greater than of positive information. A negativity bias, that is a tendency to give more weight to negative than to positive information, may be responsible for this result. The negativity bias has been demonstrated to occur in a diversity of judgment tasks (Rozin & Royzman, 2001). In addition, negative words are also detected faster, as was demonstrated in a study in which positive and negative stimuli were presented subliminally (Dijksterhuis & Aarts, 2003). Another reason can be found in the concept of self-regulation (Bandura, 1989). Theories of self-regulation are founded on a negative feedback control model. This type of system functions as a motivator and regulator of action through a discrepancy reduction mechanism. If performance matches the internal standard the person does nothing. A perceived discrepancy between performance and an internal standard will trigger action to reduce the incongruity.

Finally, we conjectured that the sensitivity to negative social feedback may lead to less automatic responses from the receiving person and activate a more critical reaction to the source of the feedback. In particular, we suggest that an actor who is perceived as capable of making sensible judgments and choices of its own, that is having high perceived agency, will be more persuasive than an actor with low perceived agency.

Figure 2: iCat

To study the influence of social feedback by a smart computer agent on user behavior we set up an experiment in which subjects received social feedback from a robotic agent while carrying out tasks in which they could conserve energy. More specifically, we tested the effects of social feedback compared to factual feedback, the effect of positive vs. negative feedback and finally we explored the effect of low vs. high perceived agency as a characteristic of the feedback source.

2 METHOD

Participants and design. Thirty-three participants (27 men and 6 women) were randomly assigned to one of three experimental conditions: a factual feedback condition, a low-agency social feedback condition, and a high-agency social

feedback condition. In addition, each participant completed 10 washing trials, which composed the 10 levels of our second independent variable. All participants were native Dutch speakers. The experiments lasted 30 minutes, for which participants were paid 3 Euros (approximately $3.75 U.S. at the time this study was conducted).

Materials and procedure. Participants were invited to engage in an experiment using a simulated washing machine. Upon arrival, they were seated individually in a small room in front of a computer. For participants in both social feedback conditions, an iCat (see Figure 2) was positioned on the participants' desk, to the right of the computer. An iCat is a robot developed by the Philips corporation[3] in the form of the stylized head of a cat, that is able, among others, to display social expressions by moving lips, eyes, eye-lashes, and eye-brows, and by playing speech files. For participants in the factual feedback condition the iCat was not present in the room. For all participants, a simulated washing machine panel was presented in the top half of the screen (see Figure 1). This panel was a copy of a current state-of-the-art model with the usual button and dials and corresponding (short) explaining words. For participants in the factual feedback condition, this panel was equipped with an energy meter that provided participants with kWh feedback about the wash program the chose. This energy meter displayed the relative amount of electricity in kWh corresponding to the chosen washing program. Similar to the McCalley & Midden study, it did not display numerical information about the level of kWh consumed. Technically, the feedback bar could differentiate between six consumption levels, which ensured that the level of detail of the feedback information was comparable between conditions. In both social feedback conditions, the energy meter was not presented. For all participants, in the bottom half of the screen, a program displaying the instructions, tasks and questions was presented. This program started with general introductions, and then instructed participants about the task: they were asked to "do washes as you do at home" by completing several simulated washing trials on a computer. Next, participants were instructed on how to program the washing machine. Participants then were given two goals: First, they were instructed to do each washing as good as possible, that is, to clean the clothes and not damage them. Second, participants were instructed to use as little electricity as possible. They were informed that washing costs electricity and that during the washing trials they would receive feedback on how much electricity the washing program they used would consume.

To participants in the factual feedback condition the program next described the energy meter and the way it provided factual feedback about the actual electricity use of the washing program they chose. To participants in the low-agency social feedback conditions, the program told that this washing machine was equipped with an "advanced electronic device" (the iCat) that the washing machine could use to give participants feedback on the amount of electricity used. In these instructions, it was underlined that the washing machine was the cause and director of all expressions presented on the "device". Next, the social feedback that the washing machine could present through the "device" was described. Participants were told that this social feedback could consist of facial expressions, little lights flashing, and short speech utterances. Some examples of positive feedback and negative feedback were described. To participants in the high-agency social feedback condition, the program described the iCat, named "Victor" and its characteristics. They were told that Victor was a very advanced robot that had a little mind of its own. They were instructed that Victor would be informing them about the energy consumption of the washing program they chose by watching their settings on screen. After one practice trial, the two goals (clean laundry and low electricity consumption) were repeated, after which the actual 10 washing trials started. For each trial (and also the practice trial), participants were instructed to complete a specific type of wash (e.g., "wash 4 very dirty jeans"). Each description of a specific type of wash was randomly drawn from a collection of thirty descriptions of common washes, for each trial of each participant such that each participant completed ten different washes. During each washing trial, participants were able to change settings on the washing machine panel until they were satisfied and were asked to press a "start" button to finalize their chosen settings. Participants received feedback about the energy consumption of the chosen washing program after each change they made. Participants in the factual feedback condition received factual feedback through the energy meter. Likewise, participants in both social feedback conditions received social feedback through the iCat during each trial. More specifically, when energy consumption was below the middle of the scale, the iCat gave positive feedback, and negative social feedback when above the middle of the scale. For small deviations from the middle of the scale, the iCat showed one of six different positive respectively one of six negative facial expressions (feedback level ‘low’, e.g., a smiling face or a sad face), for more grave deviations, the iCat illuminated little lights at the top of its ears (feedback level ‘medium’, green for positive feedback and red for negative feedback), and for the most grave deviations the iCat uttered a positive or negative word (feedback level ‘high’, e.g., "Fantastic!" or "Gruesome!"). In addition to the manipulation of agency by means of the instructions describing the iCat (described earlier), there was another difference between the low-agency and high-agency social feedback conditions. That is, in the low-agency social feedback condition the iCat used only one specific short speech utterance repeatedly to give positive social feedback during the washing trials, and another to give negative social feedback. In contrast, the iCat in the high-agency social feedback condition used a different one of six different short speech utterances to give positive social feedback, and six others for negative social feedback.

After all 10 washing trials had been completed, participants answered several demographical questions, were debriefed and thanked for their participation.

3 RESULTS

For each of the ten washing trials of each participant, we calculated the difference between the amount of electricity a participant's settings would have used, and the average usage of electricity for that specific type of wash (e.g., the 4 very dirty jeans) by all participants in our study. We labeled this

[3] see www.research.philips.com/robotics

the energy consumption score. This way, we were able to calculate a DV that indicated the difference between a reference amount of electricity needed for a specific type of wash (at least in the current study) and the electricity a participant chose to use.

This energy consumption score was submitted to a 3 (feedback condition: factual feedback vs. low-agency social feedback vs. high-agency social feedback) x 10 (washing trials: 1 to 10) MANOVA, with the last variable within-subjects. Our expectancies were confirmed. That is, we found that participants who received social feedback consumed less electricity than participants who received factual feedback, indicated by a main effect of feedback condition, $F(2, 30) = 3.42$, $p < .05$. More specifically, participants who received low-agency social feedback ($M = -.02$, $SD = .16$) and participants who received high-agency social feedback ($M = -.05$, $SD = .17$) used less electricity than participants who received factual feedback ($M = .11$, $SD = .11$), $F(1, 30) = 6.64$, $p < .05$. Note that these are energy consumption scores—therefore this indicates that factual feedback leads to more electricity consumption than the average usage for a specific type of wash,

Importantly, the electricity consumption of participants who received low-agency showed no difference from the electricity consumption of participants who received high-agency social feedback, $F < 1$.

Furthermore, we found no effect on electricity consumption by washing trial condition, nor an interaction, both F's < 1.

To be able to distinguish the effects of positive and negative feedback we calculated an index based on total number of actions of users in the user interface. This means that we not only included the final choices per trial, but all the preceding programming choices. As explained in the method section, these were all followed by (factual or social) feedback, either positive or negative. The index subtracted for each action the following choice, in terms of energy consumption effect, from the current choice, thereby indicating whether the feedback resulted in a higher or lower energy consumption score for the next following choice.

These scores were submitted to a 3 (feedback condition: factual feedback vs. low-agency social feedback vs. high-agency social feedback) x 2 (type of feedback: positive vs. negative) x 3 (feedback level: low vs. medium vs. high) MANOVA.

A main effect was found of type of feedback, indicating a stronger conservation effect of negative feedback ($M = -.21$, $SD = .40$) compared to positive feedback ($M = .10$, $SD = .39$), $F(1, 786) = 164.51$, $p < .001$. A second main effect was found of feedback level, $F(2, 786) = 10.13$, $p < .001$, indicating the strongest effect on behavior at the high feedback level ($M = -.29$, $SD = .59$) and weaker effects at the low ($M = -.02$, $SD = .34$) and medium ($M = -.02$, $SD = .31$) levels. Furthermore we found two significant 2-way interactions: condition x type of feedback ($F(2, 786) = 4.60$, $p = 0.01$); feedback level x type of feedback ($F(2, 786) = 47.04$, $p<0.0001$) and a significant 3-way interaction (condition x type of feedback x feedback level; $F(4, 786) = 4.15$, $p = 0.002$). More specifically, these findings showed that the greatest changes to conservation behavior were achieved after negative feedback (compared to positive feedback), in particular at the highest feedback level. As expected this effect was strongest for the participants who received social feedback. Surprisingly this effect was stronger in the social feedback condition with low agency compared to the social feedback condition with high agency.

4 CONCLUSIONS AND DISCUSSION

In general our findings demonstrate that social feedback as provided by an embodied agent can create behavior change among human users. Remarkably, our findings suggest that this effect is greater compared to the effects of factual feedback, which was also provided interactively. The latter type of feedback can be considered as one of the most successful types of feedback employed earlier to induce energy conservation behavior (Midden et al., 2007).

In contrast to earlier work that focused on praise, our feedback intervention included both positive and negative feedback. The effect of social feedback on energy conservation should therefore be considered as the joint effect of both positive as well as negative feedback. Interestingly however, our analysis on the level of single programming acts within trials, suggested that the direct influence of negative feedback on following programming choices was greater than the effect of the positive feedback. Moreover, it seemed as if the positive feedback encouraged participants to enhance their energy consumption. Thus, while people may feel pleased by a system that offers compliments to a user, this does not ensure that this user may also change his or her *behavior*. In particular, this may hold if a user has other goals to meet, like in our case getting clean laundry. This tentative interpretation of our findings seems to be consistent with Bandura's (1989) notion of self-regulation. A follow-up study that disentangles positive and negative feedback more firmly would be desirable to corroborate this conclusion.

One could argue that the difference between social and factual feedback might also be explained as a result of different levels of precision or conciseness of the two types of feedback (see e.g. Carenini & Moore, 2006). In our setup however, the level of detail between factual and social feedback was more or less comparable. The factual feedback was indicated in kWh , but on a scale that distinguished six consumption levels. The social feedback had a similar level of precision. It also comprised six levels that indicated the positiveness or negativeness of the message.

To our surprise, we did not find the expected effects of perceived agency. The participants in the low agency group conserved overall as much energy as those in the high agency group. Our analysis at the level of the single programming acts revealed a combined effect of feedback type and condition, suggesting that the greatest effects on user choices occurred in the social feedback group with low perceived agency. This was true for the changes in the direction of more conservation, often following negative feedback, but also for the changes to higher consumption, often resulting from positive feedback. Using our present data we cannot fully disentangle these effects. However, our explorative manipulation may have played a role here. The manipulation was to a large extent a verbal one, introducing the iCat Victor either as a robot with a 'little mind of its own' or as a mere extension of the washing machine. Our tentative explanation is that this verbal manipulation was overruled by the direct

experiences of the participants in their interactions with the iCat, which might even have caused a surprise effect among the participants in the low agency condition.

Our present findings enhance our understanding of the persuasive potential at the behavioral level of embodied agents. They also suggest issues for further exploration such as the differential effects of positive and negative feedback, either single or combined, and the influence of perceived agency, for which further conceptualization and improved measures will be needed as well.

REFERENCES

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology, 25*, 273-291.

Bandura, A. (1989). Human Agency in Social Cognitive Theory. *American Psychologist, 44,* 9, 1175-1184.

Bandura, A., & McDonald, F. J. (1963). Influence of social reinforcement and the behavior of models in shaping children's moral judgments. *Journal of Abnormal Psychology, 67,* 274-281.

Carenini, G., & Moore, J. D. (2006). Generating and evaluating evaluative arguments. *Artificial Intelligence, 170*, 925–952.

Dijksterhuis, A., & Aarts, H.A. M. (2003). On wildebeests and humans: the preferential detection of negative stimuli. *Psychological Science 14*, 1, 14–18.

Fogg, B. J., & Nass, C. I. (1997). Silicon sycophants: the effects of computers that flatter. *International Journal of Human Computer Studies, 46*, 551-561.

McCalley, L. T. (2006). From motivation and cognition theories to everyday applications and back again: The case of product-integrated information and feedback. *Energy Policy, 34*, 129-137.

McCalley, L. T., & Midden, C. J. H. (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology, 23*, 589-603.

Midden, C. J. H., Kaiser, F. G., & Mccalley, L. T. (2007). Technology's four roles in understanding individuals' conservation of natural resources. *Journal of Social Issues*, *63,* 1, 155-174.

Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media as real people and places.* Cambridge, England: Cambridge University Press/CSLI.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review, 5,* 296-320.

Wright, D. (1968). Social reinforcement and maze learning in children. *Child Development*, *39*, 1,177-183.

Emotional And Non Emotional Persuasion Strength

Irene Mazzotta[1], **Vincenzo Silvestri**[1] **and Fiorella de Rosis**[1]

Abstract. Emotional argumentation is the result of a rational form of reasoning by the Persuader, subsequently translated into messages by applying some variants of the 'classical' argumentation schemes. In this paper, we discuss which may be assumed to be the relative strength of emotional vs. rational persuasion strategies and how this has been represented in a dialogic persuasion testbed.

1 INTRODUCTION

In a dialog perspective, argumentation, rather then a predefined, integrated set of propositions, is seen as a sequence of moves in which two parties (a Persuader and a Receiver) are reasoning together on some argument. The dialogue may be more or less symmetrical, as far as the initiative in persuasion and argumentation is concerned: therefore the role of Persuader and Receiver may be fixed, or may alternate during interaction. The theory of argumentation dialogues originates from research about expert systems, that were aimed at suggesting the appropriate therapy in a given situation [1]. A key function of these systems was to support their suggestion with explanations and clarifications after requests of various kinds from the user, including critiques to the suggested plan: they therefore set the framework for subsequent developments to the problem of criticizing argumentation attempts [2]. In the multiagent system domain, this kind of dialogues were subsequently employed, by agents, to distribute and contract roles and tasks [3].

Dialogic persuasion is not restricted to dialogues in which two parties are trying to resolve a conflict of opinions or attempt to influence another participant's behaviour. Some argumentative exchanges may occur in almost any kind of context: one of the most recent examples is the case of *Online Dispute Resolution*, in which an arbitration environment supports communication and discussion in web-based groups [4, 5].

While monologic persuasion is characterized by the three steps of planning, plan revision and surface realisation that are common to any NLP task, in the 'pure' persuasion dialogues that we consider in this paper the sequence of exchanges includes some typical phases, and forms of reasoning, by the Persuader:

1. *Make a proposal*: after reasoning on the Receivers' mind (system of belies, values, goals, etc), propose some action or some claim, by giving reasons as grounds for supporting the proposal,
2. *Observe the Receivers' reaction*: what does he or she say, or express differently,
3. *Classify it* (is it a request of justification, an objection, with or without counter-argumentation, a refusal, …),
4. *Reason* (again) *on the Receivers' mind* to interpret the Receiver's reaction by placing it into her presumed set of attitudes: this requires a belief-desire-intention model of mind and reasoning [3], eventually enhanced with emotions in a BDI&E model [6],
5. *Justify it* or *defend the own proposal* if possible; *retract it* if needed, *find an alternative* and *relate new argumentation to the previous one*.

A proposal may be criticized by the Receiver in several ways: i) by questioning the goal premises, ii) by attacking them with counter-arguments alleging that one or more of them is false; iii) by undercutting the inferential link between premises and conclusion with critical questions; iv) by rebutting the practical reasoning inference with counter-arguments asserting that the conclusion is false or v) by putting forward a proposal arguing for a different action, and contending that the arguments for this opposed proposal are stronger. [2]. The Persuader must be able to respond appropriately to all these situations.

Walton's distinguishes, in the argumentation process, a first phase of 'reasoning' from a phase of 'argumentation' [7]: in the first one, the persuader reasons on the Receiver's mind to select an appropriate strategy, while in the second one this strategy is translated into a coherent message. The complexity of this process increases when argumentation becomes dialogic. At every dialogue step, the Persuader must decide which part of its reasoning to make explicit in generating the argument and which one to hide or to postpone. In addition, a refined ability to 'observe the Receiver's reaction', interpret it and reason on the consequences of this reaction on the persuasion plan must be added to the system. This new reasoning ability becomes quite complex when context, personality and emotional factors are considered: research about consumers' behaviours and attitudes contributed considerably to increasing knowledge in this domain. It is well known that determinants of effectiveness of a persuasion attempt are not only the message features, but also the source and the Receiver's features [8]. Source features are not absolute, but relative to the Receiver: a source may be more or less 'credible', 'likable', similar, 'attractive' to different Receivers. According to the Persuasion Knowledge Model, consumers recognize and evaluate persuasion attempts (and select best responses during interactions) based on the perceived effectiveness and appropriateness of the persuasion tactics rather than using product knowledge [9]. On the other hand (and maybe also because of this) Receivers may be biased towards a persuasion attempt, being skeptical, defensive or hostile, either in general or towards a particular Persuader [10]. This kind of 'resistance' to persuasion influences the Receiver's response to persuasion attempts, which may include, in the three cases, different mixtures of rational and emotional components [11].

More in general, evaluation of persuasion attempts by Receivers may be influenced by affective factors. Some of these factors are stable (like personality traits: self-esteem, self-monitoring, sensation-seeking etc), others are more or less

[1] Dept. of Informatics, Univ. of Bari, via Orabona, 4, 70126, Bari, Italy. Email: {mazzotta, silvestri, derosis}@di.uniba.it.

transient. For instance, positive mood seems to reduce systematic processing of information, whereas negative mood would enhance it [12]; positive feelings lead to more positive evaluation of information received, while the opposite seems to hold for negative feelings [13]. And finally, if a persuasion move aims at influencing the Receiver's attitude, it has been demonstrated that the Persuader's attitudes are influenced, in their turn, by the success or failure of their persuasion attempts [14].

2 RELATED WORK

Although, as we have seen, theoretical aspects of dialogic persuasion have been extensively investigated in the philosophical and the marketing studies domains, examples of dialogical persuasion prototype systems are few and quite recent. NAG (Nice Argument Generator) is a precursor of argumentation systems: it includes not only a generation component [15] but also a module aimed at interpreting the Receivers' reaction [16]. DAPHNE was the first system in which adaptation of arguments to the Receiver's 'values' was considered [17]. In ASD, Reed and Walton [18] use the language of formal dialectics to define a dialectical system in terms of *Locution rules (*statements, withdrawals, questions, challenges and critical attacks), *Commitment rules* (effects of locution rules on the two interlocutor's knowledge) and *Dialogue rules (*sequencing of communicative acts). By seeing monologues as 'inner-dialogues', Kibble [19] studied the kinds of communicative acts that are employed in persuasion dialogues and how they may be represented in rhetorical structures. Magtalo (Multiagent Argumentation, Logic, and Opinion) is a prototype environment for debate. It supports flexible intuitive interaction with data in complex debate domains to facilitate understanding, assimilation and structured knowledge elicitation. [20].

3 PORTIA

This persuasion dialogue toolbox was built after Miceli et al's [21] theory of emotional persuasion, to enable testing this theory and the methods it requires to be applied in specific domains. The prototype implements Walton's idea of separation between a 'reasoning' and an 'argumentation' phase [7], by representing with bayesian networks (BNs) the uncertainty inherent in this form of reasoning. Argumentation schemes are associated with bayesian networks in the form of xml files: they are chained-back to translate the selected strategy into Recipient-adapted messages. Answers to the user reactions to persuasion attempts are produced after reasoning on the same knowledge base.

While we refer to other papers for a description of the principles behind this prototype [22], in this paper we wish to discuss, in particular, the following aspects:

- how hypotheses about the effectiveness of alternative persuasion strategies (from psychological theories and from results of experiments) reflect, in particular, into assignment of parameters to the model;
- how knowledge representation enables comparing the effectiveness of alternative persuasion strategies (either rational, or emotional, or mixed) in a particular context, by reasoning in a "*what-if" mode;*
- how the same knowledge representation enables, as well, reacting to various kinds of users' reactions to persuasion attempts.

Although this prototype tool is domain-independent, all examples in the following sections will be taken from the healthy eating domain, which is the application area we considered so far, both in the preliminary experiments and in the theory formulation and in the building of the knowledge base on which the system was tested.

3.1 Preliminary notations

Let us introduce the following notations (synthesized in Table 1):

- a is a variable denoting an action (e.g.: 'to eat vegetables'); e_i, e_j, ... are variables denoting emotions (e.g. 'shame', 'pride', 'good mood', ...); g_i, g_j, g_h,... are formulae denoting states of the world - in particular, of R - (e.g.: 'R is in good health', 'R is in shape', 'R is overweight', but also 'R saves face', ...); the formula Feel (R,e) denotes, in particular, the affective state 'R feels the emotion e'.

- Bel, Int, A-Goal, V-Goal are modal operators that denote the various aspects of the mental state of R which are relevant in the persuasion process: that is, respectively, beliefs, intentions, active-goals and valued-goals. The first term of these operators denotes an agent, the second one is a formula. In particular:

- (V-Goal R g_i,) stands for "g_i, is a valued goal to R"; (A-Goal R g_i,) for "R's goal g_i, is active"; (Bel R (Do(R,a)$\rightarrow g_i$)) for "R believes that doing a implies g_i in a more or less near future"; (Bel R CanDo(R,a)) for "R believes that conditions hold for him to do a"; (Int R Do(R,a)) for "R intends to do a".

- The symbol '$\rightarrow$?' denotes an 'uncertain implication' and is represented in the BN with oriented arcs linking premises to conclusions. In the bayesian formalism, rule (A1$\wedge$A2 $\wedge$...$\wedge$ An) $\rightarrow$? B is interpreted as a conditional probability expression P(B|A1,A2,..,An) =m. This uncertain implication is specified with a table of the probabilities that B is true, conditional on all combinations of values for A1, A2, ..., An. It enables assigning to the premises different weights in establishing the truth value of the consequence.

- The generic strategy of *induction of intentions* is represented by the following relation:

$$[(\text{V-Goal R } g_i,) \wedge (\text{A-Goal R } g_i,) \wedge (\text{Bel R } (\text{Do(R,a)} \rightarrow g_i)) \wedge (\text{Bel R CanDo(R,a)})] \rightarrow ? \text{ (Int R Do(R,a))} \quad [i]$$

Formula	Meaning
(V-Goal R g_i)	g_i is a valued goal to R
(A-Goal R g_i)	g_i is an active goal to R
(Bel R (Do(R,a)$\rightarrow g_i$))	R believe that performing a implies achieving g
(Bel R CanDo(R,a))	R believes that he or she is in the condition to perform a
(Int R Do(R,a))	R intends to perform a
Feel(R,e_i)	R feels the emotion e_i

Table 1. Some notations

3.2 Persuasion strategies

According to Miceli et al [21], persuasion may employ a combination of 'rational' and 'emotional' arguments (see Table 2). In particular, emotions may be introduced in the persuasion process in two forms:

- by selecting an 'emotional goal' g_j (*appeal to the goal to feel an emotion* e_j *):* g_j =Feel (R,e_j). For example "To feel in good mood" and
- by activating, through the activation of an emotion e_j, an 'intermediate' goal' g_h which is instrumental to the final one g_i (*emotional activation of goals)*:

  (Bel R g_j) →? Feel (R,e_j) [ii]

  Feel (R,e_j) →? (A-Goal R g_h) [iii]

  [(A-Goal R g_h) ∧ (Bel R (g_i → g_h))] →? (A-Goal R g_i) [iv]

  For example: "*You look so overweight! Too bad...*"; e_j = shame; g_h = to save face; g_i to be in shape.

Selecting an appropriate strategy (either rational, or emotional, or a combination of the two) in a given context is a 'rational' planning task, based on some information about the Receivers. In PORTIA this information (the Receivers model) is inferred, with some level of uncertainty, from knowledge of their personality traits and living habits.

For example: *Extraverts tend to enjoy being with people, to be skilled in handling social situations and make friends easily*:

Extraverted(R) →? [EnjoyWithPeople(R) ∧ SkilledInSocialSituations(R) ∧ MakesFriendsEasily(R)]

Making friends is likely to be important to these subjects:

Extraverted(R) →? (V-Goal R MakeFriends(R))

Goals can be inferred as well, from knowledge of the user habits. An example:

[MakeSport(R) ∧ MakeCheckUps(R) ∧ LookAtTv(R)] →? (V-Goal R GoodHealth(R))

Individuals who make sport regularly, undergo regular check-ups and are interested in medical TV programs are probably interested in being in good health.

[(V-Goal R g_j) ∧ (A-Goal R g_j) ∧ (Bel R (Do(R,a)→g_i)) ∧ (Bel R CanDo(R,a))] →? (Int R Do(R,a))	
Rational strategies	
g_j is a 'rational' goal; it may be activated either emotionally (see below) or rationally, by inducing some belief that activates it.	
Emotional strategies	
Appeal to the goal to feel an emotion	g_j is an emotional goal
Emotional goal activation	an emotion e_j is activated, which in its turn activates an 'intermediate' goal' g_h which is instrumental to the final one g_j

Table 2. Rational and emotional persuasion strategies

3.3 Knowledge representation

Persuasion strategies are represented as belief networks: every uncertain implication introduced in the previous Section, instantiated with appropriate values of a, g, e_j, corresponds to an 'elementary' belief network (EBN). Other EBNs represent inferences the system is able to make about (Bel R (Do(R,a)→g_i)) or (Bel R CanDo(R,a)).

For example:

[HasFreeTime(R) ∧ LikesCooking(R) ∧ AvailableVegs(R)] →? (Bel R CanDo(R,EatVeg)) represents the statement:

Individuals who have some time free during the day, like cooking and live in a place in which good vegetables are available are probably in the condition to eat vegetables.

An *elementary argumentation plan* is associated with every EBN: this represents how that fragment of persuasion strategy can be translated into a natural language message.

In this knowledge representation, two problems occur: how to assign parameters to the EBNs and which part of that knowledge to represent in the associated argumentation plan. Let us briefly discuss the two problems.

a. Assigning parameters to belief networks

The problem of how to estimate parameters when building probabilistic models is a matter of discussion. BN parameters can be estimated by learning them from a corpus of data (frequentist approach) or according to subjective experience or common sense (neo-bayesian approach). In PORTIA, we adopted a neo-bayesian approach, by extracting knowledge on one hand from psychological theories and on the other hand from the results of our preliminary experiments [23]. In particular, in the representation of persuasion strategies and user models, the following questions are risen:

1. Which is (in [i]) *the relative impact of the various components* ((V-Goal R g_j), (A-Goal R g_j), (Bel R (Do(R,a)→g_i)) and (Bel CanDo(R,a)), with their combination of truth values) *on the intention* to perform the action? Does this impact depend on the particular type of goal? Our hypothesis is that, given a probability distribution of values for the variables in its premises, the probability of the intention-node (Int R Do(R,a)) does not depend on the goal involved, at least in the considered application domain; therefore, parameters in the EBNs that represent instances of [i] are all the same.
2. *Which are the prior – posterior probabilities of the various goals for a given user*? That is: what can we presume to be the weight of these goals in absence of any evidence, and how does this weight change, when some evidence about the user is available? As far as *goal or needs hierarchy* is concerned, we referred to [24] (see figure 1):
   - *Physiological* needs are the need to breathe, to regulate body temperature, the need for water, for sleep, the need to eat and to dispose of bodily wastes. Sexual activity is also placed in this category, as well as bodily comfort, activity, exercise etc.
   - *Safety* needs include security of employment, of revenues and resources, physical security, moral and physiological security, familiar security and health.

- *Love/belonging* needs involve emotionally-based relationships in general, such as friendship, sexual intimacy, and having a family.
- *Status* needs are the need to be respected, to self-respect and to respect others.
- *Being* needs include *self-actualization* (personal potential, self-fulfillment, seeking personal growth and peak experiences) and *self-transcendence* (helping others to achieve self-actualisation as a way of providing a route to achieve personal growth, integration, and fulfillment.

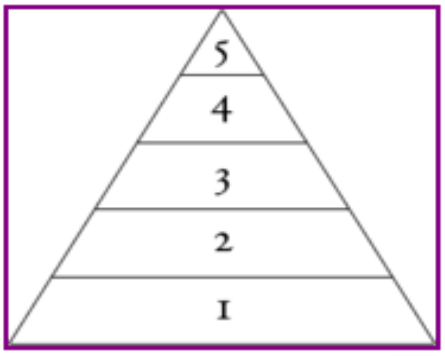

Diagram of Maslow's hierarchy of needs.
5. Actualization
4. Status (esteem)
3. Love/belonging
2. Safety
1. Physiological (biological needs)

Figure 1: Hierarchy of needs, accordingto Maslow [24]

By following this theory, we associated with higher level goals higher prior probabilities. For example: the goal of being in good health (safety in Maslow's hierarchy) has the highest weight, followed by making friends (love-belonging in the hierarchy) and having a good appearance (status-esteem). The weight of emotional goals or values, like 'to be in good mood', 'to enjoy tasting new foods' (instrumental to making friends) or 'to support farmers' (instrumental to status) are lower. Clearly, this hierarchy is only a default image of presumed goal strengths in the population: it is not identical to all individuals but can change according to specific situations. In addition, it is well known that individuals do not always behave consistently with their goals [25]: and the process of persuasion aims at re-establishing some consistency between scale of values and actual behaviour.

3. *which are the relative strengths of emotional and rational goal activation strategies*? Parameters in the EBNs representing A-goals were assigned so as to make strategies of emotional goal activation stronger than the rational ones. This was a results of our preliminary test, in which emotional strategies were considered to be more effective than rational ones and 'appeal to positive consequences' more effective than 'appeal to negative consequences' [23].
4. which are the relative strengths of alternative strategies arguing on the action-goal relation, such as Appeal to Expert Opinion or Appeal to Popular Opinion, or others? Does this strength depend on the context in which strategies are used? We suspect that the Recipient's characteristics influence the strength of strategies arguing on the action-goal relationship. For instance: 'rational' people are probably more easily persuaded by an Appeal to Expert Opinion, while very 'socialised' people might be more easily persuaded by an Appeal to a Friend's Personal Experience, ... etc. However, to our knowledge no theory or experiments supporting this hypothesis are available.

b. Building argumentation plans

Elementary argumentation plans (EAP) associated with every EBN represent how argumentation schemes [18] may be translated into message plans. Two new elements are added in this component of PORTIA's knowledge base: on one hand, hypotheses about *which items of emotional argumentation schemes should be said, and which ones should be omitted* (an instantiation of the concept of enthymeme, that is omissions of some premises that P considered in his reasoning [26]; on the other hand, definition of the rhetorical relations associated with every argumentation scheme:

- While all the components of the elementary EBNs corresponding to rational strategies are represented in the EAP, the nodes representing affective features of the Recipients (their personality traits or their emotional state) are omitted. For example, in the EAP associated with [ii, iii, iv] the activated emotion Feel(R,e_j), the instrumental goal (A-Goal R g_h) and the implication (Bel R ($g_i \rightarrow g_h$)) will not appear in the EAP.
- The following *rhetorical relations* are associated with argumentation schemes:
  (Argument from Consequences → Motivation);
  (Argument from problem to solution → Solutionhood);
  (Argument from Position To Do→ Enablement);
  (Argument from Expert Opinion or Popular Opinion or others → Evidence).

4. PORTIA AT WORK

This system is thought to be a toolbox to be used by Persuaders to receive a support in performing the tasks 1 to 5 listed in the Introduction. Although, as we said, the tool is domain-independent, to illustrate the tasks it can perform we will make some examples about healthy eating, that we selected as the application domain in this paper.

a. Selecting a 'promising' strategy by inferring the presumed strength of goals.

In this phase, the Persuader exploits its information about the Receiver to infer the presumed weight of her goals. Two kinds of information about the Receiver may be introduced into PORTIA: 'facts' about her life style (in the left side window in Figure 2) and hypotheses about her personality traits (central window). The reasoning component of PORTIA propagates this evidence into its belief networks to compute the posterior probability of the various -rational and emotional- goals (bottom window).

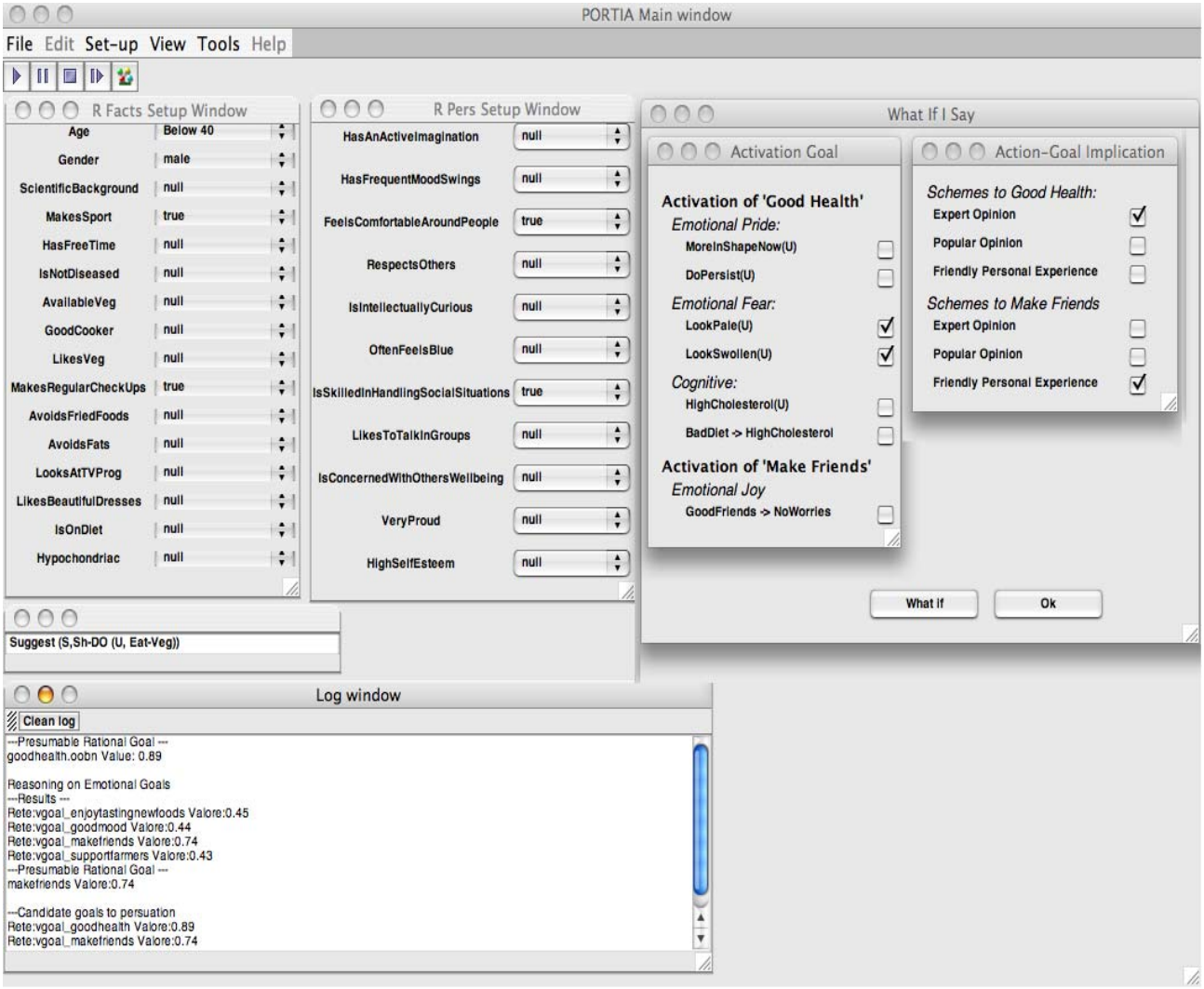

Figure 2: The interface of PORTIA for inferring the presumed goal

In the example in Figure 2, R is a man below 40 years of age who regularly makes sport and medical check-ups; he is presumed to be an *extraverted* person, as he reported to feel comfortable around people and to be skilled in handling social situations (Myers-Briggs typology questionnaire[2]). PORTIA infers, from this data, that the two candidate goals on which to support a promising persuasion strategy are *to be in good health* (rational: p=89) and *to make friends* (emotional: p=.75).

b. Selecting an appropriate goal-activation and action-goal argumentation strategy

In this phase, rather than automatically making a choice, PORTIA reasons in a 'what-if mode', to suggest alternative ways to strengthen the persuasion power of the selected strategy. Again, two types of information, expressed as sentences, may be used to influence R's attitudes: goal activation sentences and action-goal implication sentences. The first ones are focused on the (A-Goal R g), the second ones on the (Bel R (Do(R,a)→g_i)) components of implication [i]. Again according to [21], there are different ways to activate a given goal, either cognitively (by influencing beliefs which activate, in their turn, the goal) or emotionally, for instance by acting on *Pride, Shame, Emulation:* this will require instantiating appropriately the implications [ii], [iii] and [iv]. There are, as well, different ways to argue on the action–goal implication (Walton and Reed's schemes and further revisions, available at[3]): *Appeal To Expert Opinion, Appeal To Popular Opinion, Appeal To Friendly Personal Experience,* etc.

[2] http://www.humanmetrics.com/cgi-win/JTypes2.asp

[3] http://araucaria.computing.dundee.ac.uk/schemesets/walton.scm
http://araucaria.computing.dundee.ac.uk/schemesets/pollock.scm
http://araucaria.computing.dundee.ac.uk/schemesets/katzav-reed.scm

Every strategy may be triggered by one or more sentences: however, their effect depends on the Receiver's characteristics and on the context, in a way that, to our knowledge, is not yet psychologically clear. Due to this lack of background knowledge, rather than making an automatic choice on this point, alternative strategies are displayed in the 'What If I Say' window (right side of figure 2). The user (acting as a Persuader) can test the effect of alternative strategies on R's mind but is left free to make the final choice.

c. Building a dialogue plan

The dialogue plan is built by chaining-back the EAPs associated with every selected EBN. Intermediate nodes in this plan are rhetorical relations (in italic), while leaf nodes are communicative acts. Figure 3 shows an example of dialogue plan tailored to the Receiver described in Figure 2. Here, the EAP corresponding to the activation of the goal 'to be in good health' is linked with a *Solutionhood* to the rest of the plan. This includes a Suggestion, linked with a *Motivation* to the arguments in support of it. Two motivations for the suggestion are considered at the same time (and therefore in *Joint* between them): the first one is represented by a subplan for the 'rational' goal of being in good health, the second one by a subplan for the 'emotional' goal of making friends. A relation of *Evidence* links communicative acts of Inform or Remind to the claim they support. Reminds are used to mention facts that were communicated by the Receiver; Inform are used to mention facts presumed or known by the System. A relation of *Enablement* links the described part of the plan to the final subplan arguing in favour of the CanDo. The correspondence between node names inthe BN and leaf nodes in the dialogue plan is described in Table 3.

BN node name	Communicative act in the dialogue plan
(V-Goal R g_i)	Claim Like(R, g_i)
(Bel R (Do(R,a)→g_i))	Claim Implies(a, g_i)
(Bel R CanDo(R,a))	Claim CanDo(R,a)
(Int R Do(R,a))	Suggest ShDo(R,a)
Property(R)	Remind *Property*(R) or Inform *Property*(R)

Table 3. Correspondence between BN node names and communicative acts in the dialogue plan

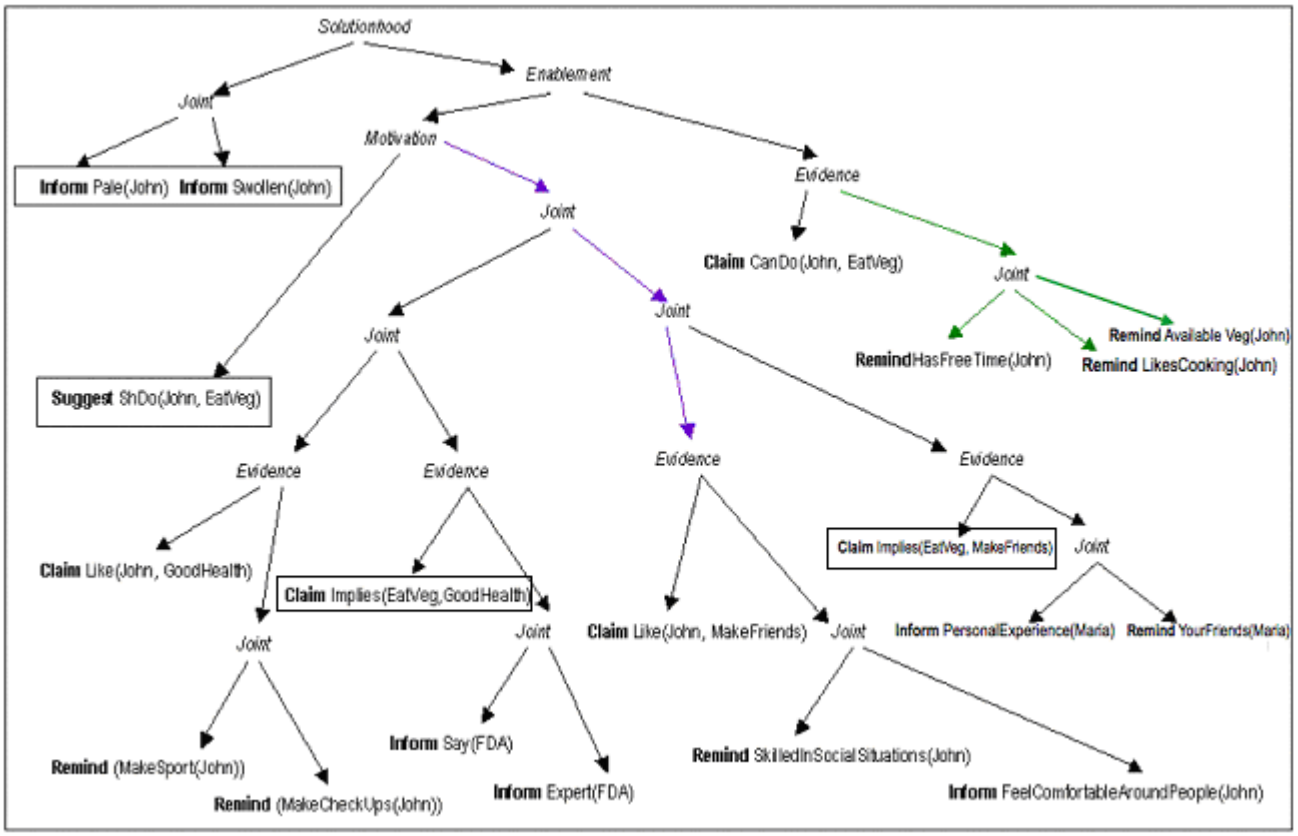

Figure 3: A dialogue plan for the example in figure 2.

d. Generating the first move

The first dialogue move is a system's Suggestion, possibly integrated with some enhancing form like an '*appeal to cognitive inconsistency*' or with a goal activation, as in the previous example. Of course, the plan in Figure 3 may be used to generate two type of Suggestion: either an enriched message in a monologic viewpoint or a simple sentence in a dialogic viewpoint.

The following system's suggestion (monologic viewpoint) can be generated from the plan in Figure 3*:*

"You look a bit pale and swollen lately, John! You should eat more fruit and vegetables, which are very good for health. In addition, a dinner rich in fresh fruit and vegetables is superb, to spend good time with friends!".

This message includes only what is considered to be the 'main' part of the plan (the framed leaf nodes in Figure 3). Other parts are omitted with the intent to avoid including too many details: therefore, the selected 'Claims' are not supported and the 'CanDo' subtree is pruned out.The last type of omissions (typical of enthymemes) regard items that are presumed to correspond to 'shared knowledge': this is the case of 'Remind' nodes which (as we said) denote information provided by the Receiver in previous phases of the dialogue, but also of subtrees arguing about the Receiver's goals (Claim-Like type of nodes and their brothers).

The plan in Figure 3 can be used, as well, to simulate a follow-up dialogue, as we will see in the next paragraphs.

e. Recognizing the Receiver's reaction

We consider two types of reactions: 'non destructive' and 'destructive' ones.

- We call '*non-destructive*' the reactions which do not involve a failure of the persuasion attempt and therefore do not require a re-planning phase. Non destructive reactions that PORTIA is able to recognize are a similar to those proposed in [19]: *RequestToJustify*, involving questioning the premises of a statement, and *Object*, alleging that a system statement is false.
- On the contrary, we call '*destructive*' the reactions that involve a failure of the attempt: these may be temporary, like a *Deny* (to deny that a goal is important to self) or permanent, like a *Rebuttal* (to claim that R is not able to perform the suggested action). permanent, like a *Rebuttal* (to claim that R is not able to perform the suggested action).

Although this is only a subset of the types of moves that can occur in persuasion dialogues, they are a good start for asymmetric dialogues, like those we are considering in PORTIA. A recognition method of the reaction type based on Latent Semantic Analysis, on which we worked with other colleagues, is sketched elsewhere [27].

f. Responding to the Receiver's reaction

A simple algorithm of exploration of the dialogue plan is applied to respond to 'non-destructive' reactions. *RequestToJustify* moves require exploring the plan tree from the identified question node, by first going to its parent -rhetorical relation node. and then down to the evidence(s) that prove it. *Object* moves require a similar plan exploration, with different kinds of answers. On the contrary, a *Deny* move is interpreted as a failure in the choice of the goal on which the persuasion strategy was focused: it requires a new reasoning and planning activity, focused on the next goal that was identified as 'promising' in the phase of *Selecting a promising strategy*; it requires, as well, a revision of the argumentation plan accordingly. Finally, a *Rebuttal* move produces a failure that cannot be repaired.

The following is an example of persuasion dialogue that can be generated from the plan in Figure 3 after the Receiver reacts in a 'non- destructive way':

S: *You look a bit pale and swollen lately, John! You should eat more fruit and vegetables.*
(Suggest ShDo (John, EatVeg))

U: *Why?*
(RequestToJustify ShDo(John, EatVeg))

S: *Because their are very good for health.*
(Claim Implies (EatVeg, GoodHealth))

U: *Yes, I know. But cooking vegetables is boring and I rather prefer spending my time with people and making new friends.*

(Object Like (John, GoodHealth)).

S: *You are right: but don't forget that a dinner rich in fresh fruit and vegetables is superb, to spend good time with friends!*

(Claim Implies(EatVeg, MakeFrieds)).

This is an ongoing part of our research.

5. CONCLUSIONS & FUTURE WORK

In this paper we described PORTIA, a persuasion dialogue toolbox based on Miceli et al's [21] theory of emotional persuasion and on Walton's idea of separation between a 'reasoning' and an 'argumentation' phase [7]. This tool enables testing this theory and the methods the theory requires to be applied in specific domains. We consider probability theory and

PORTIA is not thought to be a 'Persuader' but a *Persuasion support toolbox* for simulating persuasion dialogues. It might be used by a Persuader to compare the strength of alternative persuasion strategies, or to select the argumentation plan to follow in order to induce an *intention to change* a habit or a behaviour in a Receiver with - partially known - characteristics. For example, in the healthy eating domain PORTIA might support the Persuader to induce in the Receiver's mind the intention to contact a nutritionist, without suggesting any particular kind of diet.

PORTIA has not yet got all the potentialities of the method described in this paper. Emotional persuasion is a new research domain: the main difficulty in progressing with our work is therefore to find psychological theories on which to ground PORTIA's knowledge base. We plan to evaluate the effectiveness of our method in a near future, when PORTIA will have been completedin all its potential.

So far, we applied this tool in a domain in which we had some past experience: however, the domain-independence of PORTIA make it potentially useful in different fields, ranging from *sensitization campaigns* on medical and social aspects (like family planning or stop smoking) to *online distribution of products and e-commerce*. For example, PORTIA may be used to support interactive advertising in online shopping or telephone marketing (to subscribe telephone, energy, gas and

other contracts). In the first case, it might be integrated into an online shopping server in order to increase the user propensity towards the offered products and the communication effectiveness. In the second one, PORTIA might support the call-center operators by suggesting them a persuasive strategy to employ in their telephone work. In a far future, PORTIA might become part an embodied training agent for new call-sell operators in a virtual environment. But this is only a perspective!

ACKNOWLEDGEMENTS

This work was financed, in part, by HUMAINE, the European Human-Machine Interaction Network on Emotion (EC Contract 507422).

REFERENCES

[1] B. G. Buchanan and E. H. Shortliffe: Rule-Based Expert Systems. Addison-Wesley (1984).

[2] D.Walton: How to make and defend a proposal in a deliberation dialogue. *Artificial Intelligence & Law.* 14, 3, 177-239 (2006).

[3] M Wooldridge: An Introduction to Multiagent Systems. Chichester, Wiley (2002).

[4] G Vreeswijk and A R. Lodder: GearBi: Towards an online arbitration environment based on the design principles simplicity, awareness, orientation, and timeliness. *Artificial Intelligence And Law*, 13, 2, 297-321 (2005).

[5] D. N. Walton and D M Godden: Persuasion dialogue in online dispute resolution. *Artificial Intelligence and Law*, 13, 2 (2005).

[6] V. Carofiglio and F. de Rosis. In favour of cognitive models of emotions. *AISB Workshop on 'Mind Minding Agents'* (2005).

[7] D. N. Walton: What is Reasoning? What is Argument? The *Journal of Philosophy* 87: 399-419 (1990).

[8] D. J. O'Keefe: Persuasion: Theory and research. Thousand Oaks, CA: Sage (2002).

[9] M. Friestad and P. Wright: The persuasion knowledge model: How people cope with persuasion attempts. *The Journal of Consumer Research,* 21, 1, 1-31 (1994).

[10] R. Ahluwalia: Examination of psychological processes underlying resistance to persuasion. *Journal of Consumer Research* (2000).

[11] S. Coutinho and B. J. Sagarin: Resistance to persuasion through inductive reasoning. In *Studies in Learning, Evaluation, Innovation and Development.* 3, 2 (2006).

[12] C. R. Hullett: The impact of mood on persuasion: a meta-analysis.*Communication Research*, 32, 4, 423-442 (2005).

[13] R. E. Petty, F. Gleicher and S. M. Baker: Multiple roles for affect in persuasion. *In J P Forgas: Emotion and Social Judgements.* Pergamon Press (1991).

[14] K Geyskens, S. Dewitte and K Millet: Stimulating referral behavior may backfire for men: the effect of referral failure on susceptibility to persuasion. KUL Working Paper No. OR 0609 Available at SSRN: http://ssrn.com/abstract=944392 (2006).

[15] I. Zukerman, R. McConachy and K Korb: Using Argumentation Strategies in Automated Argument Generation. *In INLG'2000 Proceedings - the First International Natural Language Generation Conference*, 55-62 (2000).

[16] I. Zukerman, and S. George: A probabilistic approach for argument interpretation. *User Modeling and User-Adapted Interaction* 15 (1) (2005).

[17] F. Grasso, A. Cawsey and R. Jones: Dialectical argumentation to solve conflicts in advice giving. *International Journal of Human-Computer Studies*, 53, 6, 1077-1115 (2000).

[18] C. Reed and D. N. Walton: Argumentation schemes in dialogue. *Conference on Dissensus and the search for common ground* (2007).

[19] R. Kibble: Dialectical text planning. In *Proceedings of CMNA 2006.* In the scope of ECAI (2006).

[20] C. Reed, S. Wells: Dialogical Argument as an Interface to Complex Debates. *IEEE Intelligent Systems*, 22, 6 (2007).

[21] M. Miceli, F. de Rosis and I. Poggi: Emotional and non-emotional persuasion. *Applied Artificial Intelligence: an International Journal*, 20, 10, 849-880 (2006).

[22] I.Mazzotta, F.de Rosis and V.Carofiglio: A User-Adapted persuasion system in the healthy eating domain. IEEE Intelligent Systems, 22, 6 (2007).

[23] I. Mazzotta and F. de Rosis: Artifices for persuading to improve eating habits. *AAAI Spring Symposium on "Argumentation for consumers of health care"*, 76-85 (2006).

[24] A. Maslow: A theory of Human Motivation. Psychological Review, 50, 370-396 (1943).

[25] L. Festinger: A theory of cognitive dissonance. Stanford University Press, 1957.

[26] D. N. Walton: Enthymemes, Common Knowledge, and Plausible Inference. *Philosophy and Rhetoric*, vol. 34, no. 2, pp. 93-112 (2001).

[27] I.Mazzotta. N.Novielli, E.Silvestri and F.de Rosis: "O Francesca, ma che sei grulla?" Emotions and irony in persuasion dialogues. *Proceedings of the 10th Conference of AI*IA - Special Track on AI for Expressive Media'* (2007).

Social and Persuasive Argumentation over Organized Actions

Maxime Morge[1]

Abstract. To greater adoption of argumentation technologies, their links with other disciplines need attention. In particular, Sociology provides a pertinent and well-grounded background for analysing the social dimensions of multiagent organisations. In this paper, we explore the social science background which captures the notions of motivation and social power/relationship in order to provide a coordination mechanism for open complex multiagent systems. Moreover, we formalize here these notions and we apply to them a particular argumentation technology for allowing agents to negotiate. Agents argue for persuading each other to collaborate with the help of two different schemes: appeals to common goal and threats. Our framework is exemplified with a simple use case.

1 INTRODUCTION

In the past decade argumentation has become increasingly important in Artificial Intelligence. It has provided a fruitful way of approaching defeasible reasoning, decision support, dialogue, and negotiation [3]. Argumentation has been researched extensively over the last years in application domains such as law, medicine and e-democracy. [13] points out that the links between argumentation and other disciplines need attention to greater adoption of argumentation technologies. For instance, argumentation and social theory require prior theoretical development in order to develop intelligent computer systems to support collaborative work.

Sociology provides a pertinent and well-grounded background for analysing the social dimensions of multiagent organisations. In this paper, we explore the social science background which captures the notions of motivation and social power/relationship in order to provide a coordination mechanism for open complex multiagent systems. Moreover, we formalize here these notions and we apply to them the argumentation technology proposed in [12] for allowing agents to negotiate. Concretely, we formalize the *Sociology of Organized Action* (SOA) [7] through the notion of *Concrete Action Systems* (CAS). Agents argue over it for persuading each other to collaborate with the help of two different schemes: appeals to common goal and threats. Our framework is exemplified with a simple use case.

The paper is organised as follows. Section 2 presents and formalized the social theory that is the basis for our proposal, namely the SOA. The formalization of CAS focus on the major concepts of agent, resource, and goal as well as their relationships. Section 3 introduces a conceptual framework for analyzing the decision problem related to the confident behaviour of agents. Section 4 presents our computational Argumentation Framework (AF) for decision making. Section 5 outlines the social interaction amongst agents. This interaction is illustrated by two persuasion examples: the first one is using an appeal to common goal and the second one is using a threat. Section 6 discusses some related works. Section 7 concludes with some directions for future work.

[1] Universitá di Pisa, Italy, email: morge@di.unipi.it

2 FORMAL CAS

The *Sociology of Organized Action* [7] (SOA), also called *strategic analysis*, studies the interaction amongst agents within an organization with the help of Concrete Action Systems (CAS) which have been formalized in [14]. We simplify and extend here this formalization for our purpose.

In order to study the interaction amongst (human) agents durably engaged in an organization (e.g. a firm, a university, a political institution), the SOA defines *Concrete Action Systems* (CAS) as structured contexts of cooperation among agents constraining their autonomy with respect to the power relationships. The power relationships result from the mastering of one or several resources. Each actor controls some resources and needs some other resources in order to achieve their goals. Therefore, the resources are the media of the power relationships between agents. To summarize, a CAS is an analysis grid of organizations taking into account the *resources*, the *agents*, and their *goals* as well as their relationships: the agents *control* resources, the agents *need* of resources, the resources are *required* for the achievement of goals, the agents *select* goals, the goals *depends* on one another.

The concepts manipulated by a CAS can serve to study (collective) decision-making processes. For this purpose, the selection of a goal by an agent as well as the dependence of goals must be weighted. Priorities, possibly numerical weights, allow goals to be compared to one another in a rational and consistent way as envisaged by multicriteria decision making [11]. In the same way, the agent's needs (respectively controls) of resources can be weighted. An utility function (respectively payment) allows resources to be compared to one another as envisaged by the game theory [16].

Figure 1 depicts our formalization of CAS with the help of an entity-relation schema. Each resource is mastered by zero or more agents who decide about its availability, and so influence the achievement of the goals of the agents who need it. Each agent masters (needs of) zero or more resources. The agent depending on a resource affects it an *utility* to measure the satisfaction from consumption this good. The agent controlling a resource determines the *payment*, i.e. his reward for allowing its access. The agent selecting a goal affects it a *priority* to measure its importance. Moreover, the dependence relationship over goals is also associated with a priority. Obviously, the relation of requirement can be weight by a measure of necessity.

Since this weight can be calculated with the utilities, the payments, and the priorities, we do not mention it in our framework.

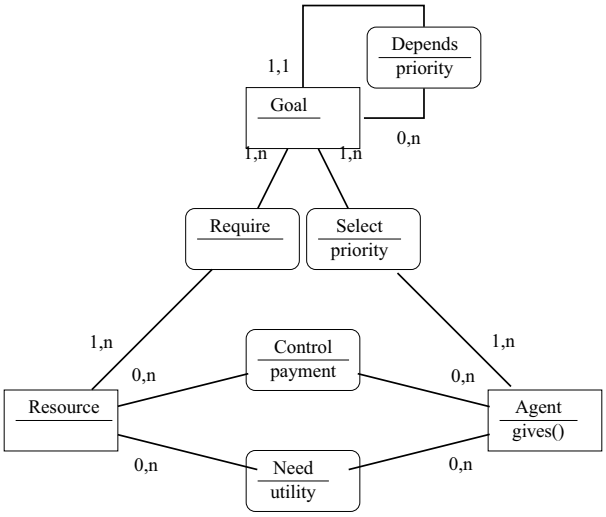

Figure 1. Formalization of a CAS

Using the convention of denoting constants in `typescript` and variables in *italics*, we illustrate the approach with the "hang a picture" [15] example, suitable adapted for illustrating persuasion. We consider here two agents `alice` and `bob`. They share the same goal which consists of seeing a picture hung in their living-room, `hung`. Each agent wants to hang the picture by oneself: `hang(alice)` or `hang(bob)`. Obviously, the fact that the picture is hung depends on one of these two goals. Moreover, `alice` is mad at `bob` and she wants to hit him, `hit(alice)`. `bob` wants to avoid it, $\neg$`hit(alice)`. We consider here two resources: a `hammer` and a `nail`. The agent `alice` (respectively `bob`) controls the `hammer` (respectively the `nail`): `control(alice, hammer)` (respectively `control(bob, nail)`). The `hammer` and the `nail` are required for hanging the picture. The `hammer` is required for hitting someone. We can deduce that `alice` (respectively `bob`) needs of the `nail` (respectively the `hammer`). For this purpose, `alice` (respectively `bob`) can give the `hammer` (respectively the `nail`) to `bob` (respectively `alice`): `give(alice, bob, hammer)` (respectively `give(bob, alice, nail)`).

3 DECISION ANALYSIS

Our methodology is to decompose the decisionproblem into elements that can be analyzed and can be brought together to create an overall representation. We use here *influence diagrams* which are simple graphical representations of decision problems [5] including the decisions to make amongst the possible courses of action (called *decision nodes*, represented by squares), the value of the specific outcomes that could result (called *value nodes,* represented by rectangles with rounded corners), and the uncertain events which are relevant information for decision making (called *chance nodes*, represented by ovals). In order to show the relationship amongst these elements, nodes are put together in a graph connected by arrows, called *arcs*. We call a node at the beginning of an arc a *predecessor* and one at the end of an arc a *successor*. The nodes are connected by arcs where predecessors are independent and affect successors. Influence diagrams which are properly constructed have no cycles. In order to capture multi-criteria decision making, it is convenient to include additional nodes (called *abstract value nodes*, represented by double line) that aggregate results from predecessor nodes. While a *concrete value* is specified for every possible combination of decisions and events that feed into this node, an abstract value is specified for every possible combination of values that feed into this node, and so the multiple attributes are represented with a hierarchy of values where the top, abstract values aggregate the lower, concrete values. We assume that influence diagrams are provided by users via a GUI which allows them to communicate user-specific preferences.

We consider here the decision problem of an agent (cf Fig. 2). The fact that the picture is hung (`hung`) depends on its decision, `give`(ag_1, ag_2, res). This top main value is split into two concrete values, the fact that an agent ag_2 hangs the picture `hang`(ag_2) and the fact that an agent ag_2 hits the other one or not (`hit`(ag_2) or $\neg$`hit`(ag_2)). The evaluation of these criteria depends on the agent knowledge, namely the information about the controls of the resources, `control`(ag_1, res). The agent also provides, through the GUI, her preferences and constraints. For instance, according to `alice`, her own goals (e.g. `hang(alice)`) have priority over `bob`'s goals (e.g. `hang(bob)`). According to `bob`, the priorities over these goals are similar.

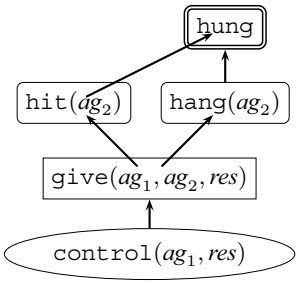

Figure 2. Influence diagram to structure the negotiation

4 ARGUMENTATION FRAMEWORK

According to the approach of defeasible argumentation of [8], arguments are reasons supporting claims which can be defeated[2] by other arguments.

Definition 1 (AF) *An* argumentation framework *is a pair* $\mathtt{AF} = \langle \mathcal{A}, \mathtt{defeats} \rangle$ *where* $\mathcal{A}$ *is a finite set of arguments and* `defeats` *is a binary relation over* $\mathcal{A}$*. We say that a set* `S` *of arguments defeats an argument* `a` *if* `a` *is defeated by at least one argument in* `S`*.*

[8] also analysis when a set of arguments is collectively justified.

Definition 2 (Semantics) *A set of arguments* $\mathtt{S} \subseteq \mathcal{A}$ *is:*

- conflict-free *iff* $\forall \mathtt{a}, \mathtt{b} \in \mathtt{S}$ *it is not the case that* `a` *defeats* `b`*;*
- admissible *iff* `S` *is conflict-free and* `S` *defeats every argument* `a` *such that* `a` *defeats some arguments in* `S`*.*
- `S` *is* preferred *iff* `S` *is maximally admissible;*

[2] The defeat relation is called attack in [8].

- $\mathtt{S}$ *is* complete *iff* $\mathtt{S}$ *is admissible and* $\mathtt{S}$ *contains all arguments* $\mathtt{a}$ *such that* $\mathtt{S}$ *attacks all attacks against* $\mathtt{a}$*;*
- $\mathtt{S}$ *is* grounded *iff* $\mathtt{S}$ *is minimally complete;*

These declarative model-theoretic *semantics* of the AF capture various degrees of justification ranging from very permissive conditions, called *credulous*, to restrictive requirements, called *sceptical*. The semantics of an admissible (or preferred) set of arguments is credulous, in that it sanctions a set of arguments as acceptable if it can successfully dispute every arguments against it, without disputing itself. However, there might be several conflicting admissible sets. That is the reason why various sceptical semantics have been proposed for the AF, notably the grounded semantics and the sceptically preferred semantics, whereby an argument is accepted if it is a member of all maximally admissible sets of arguments. For simplicity, we restrict ourself to admissible semantics.

4.1 Decision framework

Since we want to instantiate our AF for our example, we need to specify a particular framework capturing the decision problem.

Definition 3 (Decision framework) *A* decision framework *is a tuple* $\mathcal{D} = \langle \mathcal{L}, \mathcal{A}sm, \mathcal{I}, \mathcal{T}, \mathcal{P} \rangle$*, where:*

- $\mathcal{L}$ *is the* object language *which captures the statements about the decision problem;*
- $\mathcal{A}sm$*, is a set of sentences in* $\mathcal{L}$ *which are taken for granted, called* assumptions*;*
- $\mathcal{I}$ *is the* incompatibility relation*, i.e. a binary relation over atomic formulas which is asymmetric. It captures the mutual exclusion between the statements;*
- $\mathcal{T}$ *is the* theory *which gathers the statements;*
- $\mathcal{P} \subseteq \mathcal{T} \times \mathcal{T}$ *is a (partial or total) preorder over* $\mathcal{T}$*, called the* priority *relation, which captures the uncertainty of beliefs, the priority amongst goals, and the expected utilities of the decisions.*

In the object language $\mathcal{L}$, we distinguish six disjoint components:

- a set of *abstract goals* (resp. *concrete goals*), i.e. some propositional symbols which capture the abstract values (resp. concrete values) that could result;
- a set of *decisions*, i.e. some predicate symbols which capture the decision nodes;
- a set of *alternatives*, i.e. some constants symbols which capture the mutually exclusive actions for each decision;
- a set of *beliefs*, i.e. some predicate symbols which capture the chance nodes;
- the *names* of rules in $\mathcal{T}$ which are unique.

In $\mathcal{L}$, we consider strong negation (classical negation) and weak negation (negation as failure). A strong literal is an atomic first-order formula, possible preceded by strong negation $\neg$. A weak literal is a literal of the form $\sim L$, where L is a strong literal.

We explicitly distinguish *assumable* (respectively *non-assumable*) literals which can (respectively cannot) be taken for granted, meaning that they can or cannot be assumed to hold as long as there is no evidence to the contrary. Decisions (e.g. $\mathtt{give}(ag_1, ag_2, res) \in \mathcal{A}sm$) as well as some beliefs (e.g. $\mathtt{control}(\mathtt{bob}, \mathtt{nail}) \in \mathcal{A}sm$) can be taken for granted. In this way, $\mathcal{D}$ can capture incomplete knowledge.

The *incompatibility relation* captures the conflicts. We have $L \mathcal{I} \neg L$, $\neg L \mathcal{I} L$, and $L \mathcal{I} \sim L$ but we do not have $\sim L \mathcal{I} L$. We say that two sets of sentences Φ_1 and Φ_2 are incompatible ($\Phi_1 \mathcal{I} \Phi_2$) iff there is at least one sentence ϕ_1 in Φ_1 and one sentence ϕ_2 in Φ_2 such as $\phi_1 \mathcal{I} \phi_2$.

A *theory* gathers the statements about the decision problem.

Definition 4 (Theory) *A* theory $\mathcal{T}$ *is an extended logic program, i.e a finite set of rules* R*:* $L_0 \leftarrow L_1, \ldots, L_j, \sim L_{j+1}, \ldots, \sim L_n$ *with* $n \geq 0$*, each* L_i *being a strong literal in* $\mathcal{L}$*. The literal* L_0*, called the* head *of the rule, is denoted* $\mathtt{head}(R)$*. The finite set* $\{L_1, \ldots, \sim L_n\}$*, called the* body *of the rule, is denoted* $\mathtt{body}(R)$*. The body of a rule can be empty. In this case, the rule, called a* fact*, is an unconditional statement. R, called the unique* name *of the rule, is an atomic formula of* $\mathcal{L}$*. All variables occurring in a rule are implicitly universally quantified over the whole rule. A rule with variables is a scheme standing for all its ground instances.*

For simplicity, we will assume that the names of rules are neither in the body nor in the head of the rules thus avoiding self-reference problems. Considering a decision problem, we distinguish:

- *goal rules* of the form R: $G_0 \leftarrow G_1, \ldots, G_n$ with $n > 0$. Each G_i is a goal literal in $\mathcal{L}$. The head of the rule is an abstract goal (or its strong negation). According to this rule, the abstract goal is promoted (or demoted) by the goal literals in the body;
- *epistemic rules* of the form R: $B_0 \leftarrow B_1, \ldots, B_n$ with $n \geq 0$. Each B_i is a belief literal of $\mathcal{L}$. According to this rule, B_0 is true if the conditions $B_1, \ldots, B_n$ are satisfied;
- *decision rules* of the form R: $G \leftarrow D(a), B_1, \ldots, B_n$ with $n \geq 0$. The head of the rule is a concrete goal (or its strong negation). The body includes a decision literal ($D(a) \in \mathcal{L}$) and a set of belief literals possibly empty. According to this rule, the concrete goal is promoted (or demoted) by the decision $D(a)$, provided that conditions $B_1, \ldots, B_n$ are satisfied.

Due to our representation of decision problems, we assume that the elements in the body of rules are independent, the decisions do not influence the beliefs, and the decisions have no side effects.

In order to evaluate the previous statements, all relevant pieces of information should be taken into account, such as the uncertainty of knowledge, the priority between goals, or the expected utilities of the decisions. In this work, we consider that all rules are potentially defeasible and that the priorities are extra-logical and domain-specific features. We consider that the *priority* $\mathcal{P}$ which is a reflexive and transitive relation considering possible *ex æquo*. $R_1 \mathcal{P} R_2$ can be read "R_1 has priority over R_2". $R_1 \not\mathcal{P} R_2$ can be read "R_1 has no priority over R_2", either because R_1 and R_2 are *ex æquo* or because R_1 and R_2 are not comparable. The priority over concurrent rules depends on the nature of rules. Rules are *concurrent* if their heads are identical or incompatible. We define three priority relations:

- the priority over *goal rules* comes from the *preferences* overs goals. The priority of such rules corresponds to the relative importance of the combination of (sub)goals in the body as far as reaching the goal in the head is concerned;
- the priority over *epistemic rules* comes from the *uncertainty* of knowledge. The prior the rule is, the more likely the rule holds;
- the priority over *decision rules* comes from the *expected utility* of decisions. The priority of such rules corresponds to the expectation of the conditional decision in promoting/demoting the goal literal.

In order to illustrate the previous notions, let us consider the goal rules, the decision rules, and the epistemic rules from alice's viewpoint which are represented in Tab. 1. According to the goal rules,

the main goal is reached if: i) either alice hangs the picture and alice does not hit bob (cf r_{01}); ii) or bob hangs the picture and alice hits bob (cf r_{02}). alice prefers to hang the picture by herself, $r_{01} \mathcal{P} r_{02}$. According to the decision rules, the picture is hung by alice (hang(*me*)) if she controls the hammer and bob gives the nail he controls (cf r_{11}). The picture is hung by bob (hang(*you*)) if he controls the nail and alice give the hammer she controls (cf r_{12}). alice can hit bob if she controls the hammer (cf $r_{21}(ag)$). Otherwise, alice needs no resource (cf $f_{22}(ag)$). According to the epistemic rules, alice beliefs that she has the hammer, f_2.

$\mathcal{T}$

$\uparrow$ r_{01}: hung $\leftarrow$ hang(*me*), $\neg$hit(*you*)
r_{02}: hung $\leftarrow$ hang(*you*), hit(*you*)

$\mathcal{T}$

r_{11}: hang(*me*) $\leftarrow$ give(*you*, *me*, nail), control(*me*, hammer), control(*you*, nail)
r_{12}: hang(*you*) $\leftarrow$ give(*me*, *you*, *hammer*), control(*you*, nail), control(*me*, hammer)
$r_{21}(ag)$: hit(*ag*) $\leftarrow$ control(*ag*, hammer)
$f_{22}(ag)$: $\neg$hit(*ag*) $\leftarrow$

$\mathcal{T}$

f_1 : control(*me*, hammer) $\leftarrow$

Table 1. The goal rules, the decision rules, and the epistemic rules.

4.2 Arguments

Since we want that our AF not only suggests some actions but also provides an intelligible explanation of them, we adopt here the tree-like structure for arguments proposed in [17] and we extend it with suppositions on the missing information.

Definition 5 (Argument) *An* argument *built upon $\mathcal{D}$ is composed by a conclusion, a top rule, some premises, some suppositions, and some sentences. These elements are abbreviated by the corresponding prefixes. An argument a can be:*

1. *a* hypothetical argument *built upon an unconditional ground statement. If L is an assumable literal (possibly its negation), then the argument built upon a ground instance of this assumable literal is defined as follows*[3]*: $conc(a) = L$, $top(a) = \theta$, $premise(a) = \emptyset$, $supp(a) = \{L\}$, $sent(a) = \{L\}$.*
   or
2. *a* built argument *built upon a rule such that all the literals in the body are the conclusion of arguments.*

   (a) *If f is a fact in $\mathcal{T}$ (i.e. $body(f) = \emptyset$), then the* trivial *argument a built upon this fact is defined as follows: $conc(a) = head(f)$, $top(a) = f$, $premise(a) = \emptyset$, $supp(a) = \emptyset$, $sent(a) = \{head(f)\}$.*

   (b) *If r is a rule in $\mathcal{T}$ with $body(r) = \{L_1, \ldots, L_j, \sim L_{j+1}, \ldots, \sim L_n\}$ and there is a collection of arguments $\{a_1, \ldots, a_n\}$ such that, for each strong literal $L_i \in body(r)$, $conc(a_i) = L_i$ with $i \leq j$ and for each weak literal $\sim L_i \in body(r)$, $conc(a_i) = \sim L_i$ with $i > j$, we define the* tree *argument a built upon the rule r and the set $\{a_1, \ldots, a_n\}$ of arguments as follows: $conc(a) = head(r)$, $top(a) = r$, $premise(a) = body(r)$, $supp(a) = \cup_{a' \in \{a_1, \ldots, a_n\}} supp(a')$, $sent(a) = \cup_{a' \in \{a_1, \ldots, a_n\}} sent(a') \cup body(r) \cup \{head(r)\}$. The set of arguments $\{a_1, \ldots, a_n\}$ are called the set of* subarguments *of a (denoted $sbarg(a)$).*

[3] θ denotes that no literal is required.

The set of arguments built upon $\mathcal{D}$ is denoted $\mathcal{A}(\mathcal{D})$.

Notice that the subarguments of a tree argument concluding the weak literals in the body of the top rule are hypothetical arguments. Indeed, the conclusion of an hypothetical argument could be a strong or a weak literal while the conclusion of a built argument is a strong literal. As in [17], we consider composite arguments, called *tree* arguments, and atomic arguments, called *trivial* arguments. Contrary to other definitions of arguments (set of assumptions, set of rules), our definition considers that the different premises can be challenged and can be supported by subarguments. In this way, arguments are intelligible explanations. Moreover, we consider *hypothetical* arguments which are built upon missing information or a decision. In this way, our framework allows to reason further by making suppositions related to the unknown beliefs and over possible decisions.

In our example, the argument b (respectively a), concludes that the main goal is reached since bob (respectively alice) hangs the picture and alice hits (respectively does not hit) bob if we suppose that alice (respectively bob) gives the hammer (respectively the nail) and if we suppose that bob controls the nail. The arguments a is depicted in Figure 3. An argument can be represented as a tree where the root is the conclusion (represented by a triangle) directly connected to the premises (represented by losanges) if they exist, and where leafs are either some suppositions (represented by circles) or the empty set. Each plain arrow corresponds to a rule (or a fact) where the head node corresponds to the head of the rule and the tall nodes are in the body of the rule. The tree argument a is composed of one trivial subargument and one tree argument. Neither trivial arguments nor hypothetical arguments contain subarguments.

4.3 Interactions

The interactions amongst arguments may come from their conflicts, from their nature (hypothetical or built), and from the priority of rules. We examine in turn these different sources of interaction.

Since their sentences are conflicting, the arguments interact with one another. For this purpose, we define the following attack relation.

Definition 6 (Attack relation) *Let $a, b \in \mathcal{A}(\mathcal{D})$ be two arguments.* a attacks b *iff $sent(a)$ $\mathcal{I}$ $sent(b)$.*

This relation encompasses both the direct (often called *rebuttal*) attack due to the incompatibility of the conclusions, and the indirect (often called *undermining*) attack, i.e. directed to a "subconclusion". According to this definition, if an argument attacks a subargument, the whole argument is attacked.

Since arguments are more or less hypothetical, we define the size of their suppositions.

Definition 7 (Supposition size) *Let $a \in \mathcal{A}(\mathcal{D})$ be an argument. The* size of suppositions *for a, denoted $suppsize(a)$, is the number of suppositions of a: $suppsize(a) = |supp(a)|$.*

The size of suppositions for an argument is the number of decision literals and assumable belief literals in the sentences of the argument.

Since arguments have different natures (hypothetical or built) and the top rules of built arguments are more or less strong, we define the strength relation as follows.

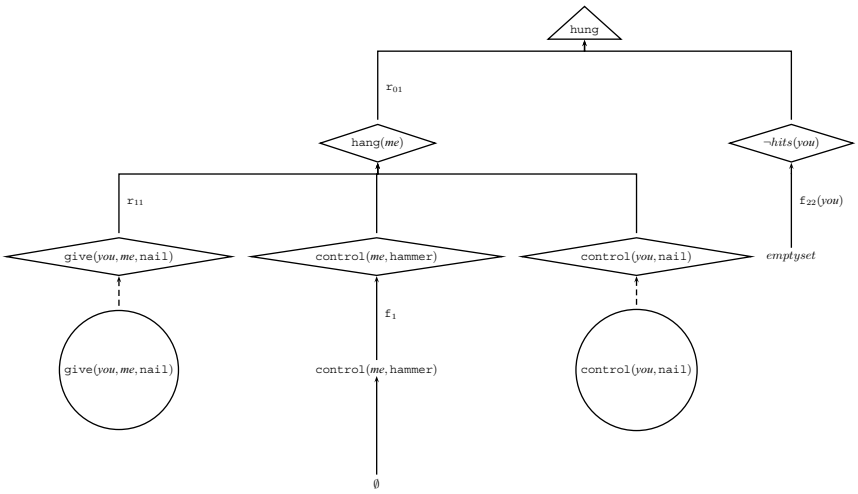

Figure 3. Arguments supporting the fact that bob give the nail

Definition 8 (Strength relation) *Let A_1 be a hypothetical argument, and A_2, A_3 be two built arguments.*

1. *A_2 is stronger than A_1 (denoted $A_2 \; \mathcal{P}^{\mathcal{A}} \; A_1$);*
2. *If $(top(A_2) \mathcal{P} top(A_3)) \wedge \neg(top(A_3) \mathcal{P} top(A_2))$, then $A_2 \; \mathcal{P}^{\mathcal{A}} \; A_3$;*
3. *If $(top(A_2) \mathcal{P} top(A_3)) \wedge (suppsize(A_2) < suppsize(A_3))$, then $A_2 \; \mathcal{P}^{\mathcal{A}} \; A_3$;*

Since $\mathcal{P}$ is a preorder on $\mathcal{T}$, $\mathcal{P}^{\mathcal{A}}$ is a preorder on $\mathcal{A}(\mathcal{T})$. Since it is preferable to consider fewer suppositions as possible, built arguments are preferred to hypothetical arguments. Moreover, we want to take into account the preferences captured by the priorities. That is the reason why we consider that an argument is stronger than another argument if the top rule of the first argument has a proper higher priority than the top rule of the second argument, or if it is not the case but the number of suppositions made in the first argument is properly smaller than the number of suppositions made in the second argument.

In order to adopt Dung's seminal calculus of opposition, we define the defeat relation.

Definition 9 (Defeat relation) *Let $a, b \in \mathcal{A}(\mathcal{D})$ be two arguments.* a defeats b *iff: i)* a attacks b*; ii)* $\neg(b \; \mathcal{P}^{\mathcal{A}} \; a)$.

Let us consider our previous example. The arguments a and b attack each other. Since the top rules of a is r_{01} and the top rule of b is r_{02}, a is stronger than b, and so a defeats b. The argument a is in an admissible set, and so can justified the opinion of alice. This arguments is useful for alice to justify its choice in front of bob and to persuade the latter.

5 SOCIAL INTERACTION

The social statements are exchanged during dialogues and notified in the *dialogical commitments*. Our agent drives the interactions by the adherence to protocols.

The negotiation is driven according to the individual/social statements concerning the goals of agents (their own goals and the goals of their interlocutors), the decisions they make, the knowledge, and preferences over them. The social statements are exchanged during dialogues and notified in the *dialogical commitments* which are internal data structures which contain propositional/action social obligations involving the agent, namely with the agent being either the debtor or the creditor. The choice amongst actions is made according to the agent's statements and the preferences over them. The dialogical commitments of alice include commitments involving alice: either alice is the creditor of the commitment, or alice is the debitor of the commitment (see the next section).

A protocol is required to conduct the interaction. For this purpose, the social reasoning uses a boot strap mechanism that initiates the required protocol, the role the agent will play in that protocol, and the other participants. The protocol engine determines the appropriate message to be sent given those parameters. When there is a decision to be made either between the choice of two locutions (e.g. an accept or a reject) to be sent or the instantiation of the content of the locution (e.g. the definition of a proposal), the protocol engine uses a precondition mechanism to prompt the social reasoning. Upon the satisfaction of the precondition, the protocol engine sends the locution. A similar mechanism is used for incoming messages. If it is necessary to update the dialogical commitments of the agent, this can be done with the post condition mechanism which operates in a similar manner.

The agents utter messages to exchange goals, decisions, and knowledge. The syntax of messages is in conformance with a common communication language. We assume that each message: has an identifier, M_k; is uttered by a speaker (S_k); is addressed to a hearer (H_k); responds to a message with identifier R_k; is characterised by a speech act A_k composed of a locution and a content. The locution is one of the following: question, assert, accept, why, withdraw (see Table 2 below for examples). The content is a triple consisting of: a goal G_k, a decision D_k, and a knowledge K_k[4]

Figure 4 depicted our protocol from the initiator viewpoint with the help of a deterministic finite-state automaton. The choice of locutions to send depends on the way the social reasoning fulfills preconditions. For example, the outcome of evaluate decision by the social reasoning will dictate to the protocol engine whether it sends accept, assert or why. The corresponding rule of the protocol engine is as follows:

[4] We will use θ to denote that no goal is given and $\emptyset$ to denote that no knowledge is provided.

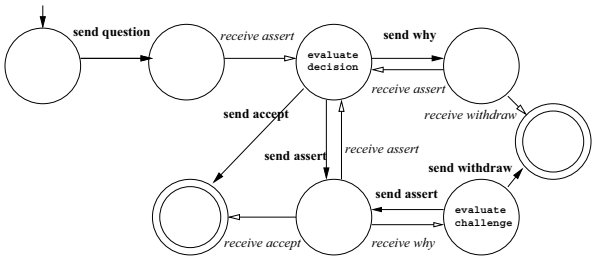

Figure 4. Protocol for the initiator

```
IF receive assert(G,D,K) from interlocutor THEN {
  update commit(interlocutor,[G,D,K]);
  IF evaluate(G,D,K) THEN{
        send accept(G,D,K) to interlocutor;
        commit(me,[G,D,K]);}
  ELSE IF evaluate(G,D2,K2) & D2!=D THEN{
              send assert(G,D2,K2) to interlocutor;
              commit(me,[G,D2,K2));}
  ELSEIF send why(G,D,K) to interlocutor;}
```

In this rule me denotes the reasoning agent and interlocutor denotes the agent it dialogues with. evaluate(G,D,K) is a predicate which evaluates if the goal G is supported by an admissible argument built upon the decision D and the knowledge K. According to the corresponding rules, the dialogical commitments are updated when a proposal is received. If an admissible proposal have been suggested, then the speech act is an accept. If a new admissible proposal is found, then the speech act is an assert. Otherwise the speech act is a why.

Table 2 depicts the speech acts exchanged between alice and bob playing two different dialogues. They attempt to come to an agreement on the transaction for the resources to reach the common goal hung. The dialogue is initiated by alice. With the message M_1, bob informs alice that it finds out that the action give(alice, bob, hammer) is justified with respect to the common goal (hung). However, alice does not find give(alice, bob, hammer) justified and she proposes give(bob, alice, nail). Since none of these proposals have been jointly accepted, bob attempts to determine the reasons for alice's choice (cf M_3). In the first dialogue (Top of Table 2), alice argues with the goal hang(alice) which is a subgoal hung. In the second dialogue (Top of Table 2), alice argues with the goal hit(alice) demoted by bob. Given alice's response in M_4, bob includes the argument provided by alice. Therefore, it finds Alice's proposal justified whatever the dialogue is. Finally, bob communicates his agreement with the help of an accept (M_5) which closes the dialogue. We can notice that the influence of alice on bob leads this latter to concede the alternative give(bob, alice, nail) which was previously not justified from bob's viewpoint. The agents are able to persuade each other, through argumentation. In the first dialogue, bob includes a new argument concluding a common goal. In the second dialogue, bob includes a new argument concluding a goal he demotes. Whatever the dialogue is, alice persuades bob.

6 RELATED WORKS

We have presented here a model of social and persuasive argumentation. In this perspective, [2] identifies different types of arguments: threats, rewards, appeals to past promise, appeals to prevailing practice, appeals to self-interest, and counter-examples. In our paper, we focus on the threats and we introduce the appeals to common goal. Actually, the concepts of threat and promise are two faces of the same coin. Both of them are based upon the power/dependence relations of agents [9]. The purpose of [2] is to handle these types of arguments within a classical argumentation-based framework. Even if our argumentation-based framework is different, our definitions of arguments are quite similar. However, the strengths of these arguments are different. While the strength of arguments of [2] is based upon the weakest link principle, the strength of our arguments is based upon the last link principle and depends on the hypothetical nature of arguments. Contrary to [2], we adopt the calculus of opposition of Dung [8]. Actually, our work consists of a more refined framework for persuasive negotiation than [15] where the strength of arguments is not only based upon some authority relations.

We have used here the argumentation-based mechanism for decision making proposed in [12]. The framework of [10, 12] incorporates abduction on missing information, while the frameworks of [1, 12] can be applied to a multi-criteria decision making. To the best of our knowledge, the framework of [12] is the only one integrating both of these proprieties required by our application. Moreover, the other existing frameworks do not come with a conceptual framework for creating a model and a representation of decision problems. By relying on [12], the decision problem is firstly analyzed, and so treated.

[14] proposes straight translations of the sociology of organized action into a computer science formalism. We have simplified and extended here this formalism for our purpose. The relation of requirement is weight by a measure of necessity in [14]. Since this weight can be calculated with the utilities, the payments, and the priorities, we do not mention it in our framework. Contrary to [14], we do not distinguish the resources and the stake they represent, i.e. the power relationships amongst the agents which either control the stake or depend on its. Moreover, we have added a relation amongst goals to reflect that they can depend on one another. This extension enrich the coordination model.

According to [6], the four main recurring social dimensions of multiagent organisations are: the social structures of roles and groups, the dialogical structures of the interaction amongst agents, the functional structures of the goals and tasks, and the normative structures incorporating the deontic notions of obligation and permission. With respect to this analysis grid, our framework treats a subset of the social structures captured by the power of agents resulting from the mastering of resources. Our framework considers the functional structures due to the goal decompositions captured by the the dependence relation. Obviously, the proposed negotiation protocol consists of the dialogical structure of our framework. Finally the normative structure of multiagent systems is out of the scope of our framework.

7 CONCLUSIONS

In this paper, we have described a model of autonomous, social, and argumentative agent trying to persuade each other to collaborate. For this purpose we have formalized a social theory that is served to study collective decision-making processes with the help of the concepts of *agent*, *resource*, and *goal*, as well as their relationships. Moreover, we have provided an AF for decision-making to perform the social reasoning which is about how to achieve the individualistic, the social, and the common goals through collaboration. Actually, we have developed a model of internal dialectics between the individual goals and the social goals of agents to capture the interactions which can

M_k	S_k	H_k	A_k	R_k
M_0	alice	bob	question(hung, give(ag_1, ag_2, res), $\emptyset$)	θ
M_1	bob	alice	assert(hung, give(alice, bob, hammer), $\emptyset$)	M_0
M_2	alice	bob	assert(hung, give(bob, alice, nail), $\emptyset$)	M_1
M_3	bob	alice	why(hung, give(bob, alice, nail), $\emptyset$)	M_2
M_4	alice	bob	assert(hang(alice), give(bob, alice, nail), [control(bob, nail)])	M_3
M_5	bob	alice	accept(hung, give(bob, alice, nail), $\emptyset$)	M_1
M_k	S_k	H_k	A_k	R_k
M_0	alice	bob	question(hung, give(ag_1, ag_2, res), $\emptyset$)	θ
M_1	bob	alice	assert(hung, give(alice, bob, hammer), $\emptyset$)	M_0
M_2	alice	bob	assert(hung, give(bob, alice, nail), $\emptyset$)	M_1
M_3	bob	alice	why(hung, give(bob, alice, nail), $\emptyset$)	M_2
M_4	alice	bob	assert($\neg$hit(alice), give(bob, alice, nail), [control(bob, nail)])	M_3
M_5	bob	alice	accept(hung, give(bob, alice, nail), $\emptyset$)	M_1

Table 2. Dialogue using an appeal to common goal (top) and using a threat (bottom)

exist between the agents interests and their social responsibilities. In order to valid this approach, we use the multiagent platform GOLEM [4] for the deployment of our agents.

The notions of social welfare can be used as a criterion to discriminate amongst arguments for the allocation of resources amongst agents taking into account the utilities/payments (resp. preferences) that agents assign to the resources (respectively goals). Existing results indicate that interaction amongst agents as well as the agents' reasoning can be designed to direct negotiation towards high quality allocations. Future works will aim at building upon the existing results to design and realise effective argumentation-based negotiation mechanisms that can be used to exhibit social-welfare related properties.

ACKNOWLEDGEMENTS

This work is supported by the Sixth Framework IST programme of the EC, under the 035200 ARGUGRID project. The author would like to thank Prof. Christophe Sibertin-Blanc for the precious comments regarding the ideas proposed in this paper.

REFERENCES

[1] Leila Amgoud, 'A general argumentation framework for inference and decision making', in *Proc. of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, ed., Tommi Jaakkola Fahiem Bacchus, pp. 26–33, (2005).

[2] Leila Amgoud and Henri Prade, 'Handling threats, rewards, and explanatory arguments in a unified setting', *International journal of intelligent systems*, **20**(12), 1195–1218, (2005).

[3] Trevor Bench-Capon and Paul Dunne, 'Argumentation in artificial intelligence', *Artificial Intelligence*, **171**(10-15), 619–641, (July-October 2007).

[4] Stefano Bromuri and Kostas Stathis, 'Situating cognitive agents in GOLEM', in *Proc. of the Engineering Environment-Mediated Multiagent Systems Conference (EEMMAS)*, eds., Danny Weyns, Sven Brueckner, and Yves Demazeau, pp. 76–93, Leuven (Belgium), (2007). Katholieke Universiteit Leuven.

[5] Robert Taylor Clemen, *Making Hard Decisions*, Duxbury. Press, 1996.

[6] L.R. Coutinho, J.S. Sichman, and O. Boissier, 'Organizational modeling dimensions in multiagent systems', in *Iberagents*, pp. 1–15, Ribeirao Preto, Brazil, (October 2006).

[7] Michel Crozier and Erhard Friedberg, *L'acteur et le système : les contraintes de l'action collective*, Seuil, 1977.

[8] Phan Minh Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artificial Intelligence*, **77**(2), 321–357, (1995).

[9] Margco Guerini and Cristiano Castelfranchi, 'Promises and threats in persuasion', in *Proc. of the ECAI Workshop on Computational Models of Natu- ral Argument*, pp. 1–8, Riva del Garda, (September 2006).

[10] Antonis Kakas and Pavlos Moraitis, 'Argumentative-based decision-making for autonomous agents', in *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 883–890. ACM Press, (2003).

[11] Ralph Keeney and Howard Raiffa, *Decisions with multiple objectives: Preferences and value tradeoffs*, J. Wiley, 1976.

[12] Maxime Morge, 'The hedgehog and the fox. an argumentation-based decision support system', in *Argumentation in Multi-Agent Systems: Fourth International Workshop ArgMAS, Revised Selected and Invited Papers*, eds., Iyad Rahwan, Simon Parsons, and Chris Reed, volume 4946 of *Lecture Notes in Artificial Intelligence*, pp. 114–131, Honolulu, Hawai, USA, (2008). Springer-Verlag.

[13] Iyad Rahwan and Peter McBurney, 'Argumentation technology', *IEEE Intelligent Systems, Special Issue on Argumentation Technology*, **22**(6), 21–23, (November/December 2007).

[14] C. Sibertin-Blanc, F. Amblard, and M. Mailliard, 'A coordination framework based on the sociology of organized action', in *Proc of AAMAS 2005 International Workshops on Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems*, volume 3913 of *Lecture Notes in Computer Science*, pp. 3–17. Springer Berlin, (2006).

[15] Carles Sierra, Nick R. Jennings, Pablo Noriega, and Simon Parsons, 'A framework for argumentation-based negotiation', in *Proc. of the 4th International Workshop on Agent Theories, Architectures, and Languages (ATAL)*, volume 1365 of *Lecture Notes in Computer Science*, pp. 177–192, Provident, RI, (1997). Springer.

[16] John von Neumann and Oskar Morgenstern, *Theory of games and economic behavior*, Princeton University Press,, Princeton, 1944.

[17] Gerard Vreeswijk, 'Abstract argumentation systems', *Artificial Intelligence*, **90**(1-2), 225–279, (1997).

MAgtALO: An Agent-Based System for Persuasive Online Interaction

Simon Wells and **Chris Reed**[1]

Abstract. This paper provides a concise introduction to the MAgtALO system. This prototype software environment provides a mechanism enabling users to engage in online debate using naturalistic dialogue underpinned by sound argumentation theory. MAgtALO is used to demonstrate how dialogue protocols can be applied to support flexible intuitive interaction within complex and contentious problem domains.

1 ARGUING ONLINE

Recently the online community has spontaneously demonstrated great interest in argument. This may have been spurred in part by the highly visible arguments with strong and explicit argumentative structure such as those found in the Iraq Study Group Report.

Various online systems have been deployed to exploit this interest such as *convinceme.net* and *debatepedia.com*. These systems have enabled members of the public to engage with each other and to express their opinions using web-based interaction mechanisms. In convinceme.net the aim is competitive, users attempt to accumulate votes, which equate to points, with the aim of accumulating the most points and thereby becoming *King of the Hill* within a single debate. However, Debatepedia uses a Wikipedia style interface to collate large evidence sets to support a user in exploring and understanding a complex debate topic. What both of these systems have in common is that they both provide high quality Web 2.0 based interfaces and environments to support user interaction. The interfaces arguably contribute to the construction of the broad userbases which underpin these systems but the underlying argumentation theory is often impoverished, having small sets of moves which the users can make and limited tools for interacting with the argumentative content.

Even where systems have not been designed to explicitly support the dialectical nature of argumentative dialogue, users will try to fit in a rudimentary yet intuitive argumentative structure. This occurs often in blog comments and in the BBC's *Have Your Say* webpage in which users will often manually copy and paste earlier posts into their own response to specify the exact point to which they are responding. Users obviously want a way to structure their interactions and responses so that they can explore the arguments of others whilst making their own explicit in relation. Such capabilities are not however explicitly supported by the current crop of weblog commenting or fora software. MAgtALO (MultiAGenT Argument, Logic and Opinion) is a word for disagreement used in the Tagalog language spoken in the Republic of the Phillipines, and disagreement is also a natural state of affairs in complex and contentious real world domains.

[1] School of Computing, University of Dundee, email: {swells,chris}@computing.dundee.ac.uk

The MAgtALO system adopts the intuitive and appealing interaction mechanisms found in extant online debate systems and marries them with cutting edge research into the representation of argument, as structures of knowledge, and argumentation, in terms of structured protocols for interaction. It has been suggested that argument provides an intuitive and accessible way to present and assimilate complext data [2], and that structured argumentation can be applied to discussion of complex domains involving real risks [5]. In Magtalo, both monologic argument structures and dialogic argument protocols are used to give the user intuitive control over navigation of a complex disagreement space. Presenting and organising material explicitly as arguments should mean that users find it easier to understand the relations between the various positions in comparison to sources which have a more discursive style (such as newspaper reports).

2 MAgtALO Architecture

MAgtALO consists of a multiagent-based backend server and a web-based user-facing interface which is shown in figure 1. The web interface is AJAX-based and incorporates client-side javascript to ensure that a responsive user interface is provided to the user.

Figure 1. The main MAgtALO interface

The interface is written in PHP and is served from a standard Apache web server, MySQL database, and Hypertext Pre-Processor AMP software stack. The interface enables a user to engage in a persuasive dialogue with a number of software agents using simple, in-

tuitive dialectical-game based interaction protocols. Given an initial topic the user is provided with various choices, such as to agree or disagree with a given position, or to provide a reason for their stated position. The agent back-end uses the Jackdaw University Development Environment (JUDE), an academic-oriented distribution of the Jackdaw Multiagent System (MAS) platform developed by Calico Jack Ltd [3]. JUDE is a Java-based, lightweight, flexible, and industrial-strength agent development platform that takes a modular, dynamically extensible approach to agent development. The agent backend supports the storage of user positions for later reuse in subsequent dialogues, and provides the mechanism whereby a user receives an appropriate response from the pre-stored arguments. The software agents respond to the points made in the dialogue, either by the user or by other agent participants as required, utilising information stored in their belief databases and a *desire to speak* function which controls how much an agent wishes to make an utterance at any given point in the dialogue.

3 ARGUMENTS & POINTS OF VIEW

MAgtALO uses agents in a multiagent system to represent the views of participants. Pre-existing arguments are analysed in terms of their argumentative structure using tools such as Araucaria [6]. Arguments are stored using AML, the XML-based Argument Markup Language, and are read into the agents respective belief databases. For a given domain, arguments from various perspectives can initially be analysed and stored as AML. Agents can then have their beliefs automatically populated not only with propositions that correspond to real, analysed natural text, but also with the argumentative relations that hold between those propostions. For MAgtALO this process involves finding several corpus arguments that are in the same domain but that have been articulated by different, possibly conflicting, authors. Given multiple points of view, garnered from specific authors who hold particular positions with respect to the domain, it is possible to interact with the data in interesting ways. Interaction in MAgtALO utilises the specific metaphor of a meeting, with the user holding the privileged position of chair. The user is therefore able to interact with specific agents, elicting their personal views about a particular point, asking for arguments *pro* or *con* a given position, asking an agent to attack the arguments of another, or whether they concur with the current claim, and so on. The user therefore has some control over the direction of the dialogue modulo the rules of the interaction protocol.

4 PERSUASIVE DIALOGUE

One aim of MAgtALO is to enable the participants to engage in a, possibly persuasive, discussion rather than an interrogation. This means that the protocol by which the players interact must allow for more sophisticated behaviours than just questioning the other players, building a pool of answers associated with the participants, and thereby exploring the agent knowledge bases. Each participant, including the user(s) and any agents, must be able to interject with their own opinions, especially when something is said with which they disagree. To support this kind of behaviour, two approaches were taken. The first utilises a simple dialogue game protocol that was developed to govern the kinds of things that the players can say at each point in the dialogue. Dialogue games are basically turn-taking games which can be used to structure the interactions between a dialogue's participants, enabling the participants to, for example, construct a position with respect to the point at issue. Dialogue games are based on the notion that a particular class of utterances can be classified as *speech acts* [7] or *performatives* and that when uttered they have the characteristics of actions. In a dialogue game the kinds of moves that can be made, and hence the kinds of things that can be said correspond to particular speech acts. For example, the influential game, 'DC' due to Mackenzie [4] includes the statement, withdrawal, question, challenge, and resolution demand performatives. In addition to the performatives, dialogue games are commitment based meaning that when a participant makes a move it affects their commitment with respect to the content of that move. For example, in DC uttering a statement causes the speaker to become committed with respect to that statement. Dialogue games also specify a protocol for how a dialogue can legally develop, for example, specifying that after a question, the next move must be an either a statement or withdrawal. Dialogue games have been used to analyse various errors in reasoning such as the fallacy of begging the question [4] and as normative ideals for discourse in specific domains such as ethical discussion [8]. A fragment of the dialogue game protocol can be seen in figure 2 in which the user is asking the agent Martin a question. In response to the current point, the user is able to ask Martin to supply a reason for the current point, or to supply further reasons if the proffered reason is not sufficient, or Martin can be asked whether he agrees or whether he can expand on his current position.

Figure 2. Dialogical interaction in MAgtALO

The second approach was to introduce the notion of a desire-to-speak function, which is incorporated into the agent participants of the dialogue, and enables them to automatically interject after a statement is made that exceeds their desire-to-speak threshold. The function is simple and merely calculates the difference between the number of points in support of a statement and the number of points against within an agents knowledge base. Although the user is nominally in control of the dialogue, agents may automatically interject after a statement is made if the agent has a sufficiently strong desire to speak regarding that statement. The function that currently calculates desire-to-speak is simple: it is the difference between the number of points in support and the number of points against the statement within an agents knowledge base. If the value is around zero then the agent has mixed feelings regarding the point. If the value is greater or lesser than zero, then the agent has strong feelings either for or against the point. Each agent has a threshold value set which enables the strength of feeling for a given point to be determined individu-

ally. If the threshold is exceeded then the agent will automatically express its viewpoint in the dialogue at that point. Fox and Das [2] have demonstrated that very simple aggregation functions are often all that is required for appropriate automated reasoning in many situations, a mechanism such as automatic interjection can therefore make the dialogue seem much more natural. The aim of Magtalo is not to calculate a ”solution” to a debate, or to evaluate points of view, or even particularly to persuade a user that a particular viewpoint is superior but is mainly to provide a robust software environment in which complex domains can be explored. During this process however it is likely that a user, exposed to arguments that they had not previously considered might find themselves persuaded to adjust their position. Though such things may be interesting to investigate (as is hinted at, at least in part, in section 7), they are peripheral to the main focus, which is squarely upon providing a rich, flexible, but intuitive interface by which online users can interact with and explore complex debates, thereby gaining a deeper and more sophisticated understanding of the topic. One rather more direct additional benefit of using the theory of dialogue games as a foundation upon which to build such an interface is that the process of extracting structured knowledge from the user is made significantly easier.

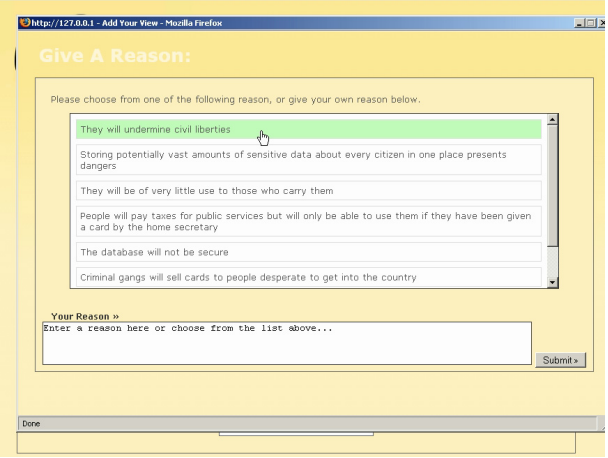

Figure 3. Knowledge elicitation in MAgtALO

5 KNOWLEDGE ELICITATION

The process of uncovering a users position on a given topic is a form of knowledge elicitation, what [8] refer to as the maieutic function of dialogue. MAgtALO uses a simple dialogue game protocol to expose this knowledge and to record it into the system in a structured fashion using AML. These AML records can subsequently be loaded into a MAgtALO Agent so that subsequent dialogues can occur in which the arguments of the current participant become the beliefs of an intelligent software agent representing the views of the original participant. Use of a dialogue game enables the underlying argumentative structure of the dialogue to be captured. This is because each statement is uttered in relation to some earlier statement. For example, offering justification for agreement with a position corresponds to an inference being drawn between the two points, one giving a conclusion and the other giving a reason in support of the conclusion. The use of a dialogue game protocol therefore ensures that each new entry into the dialogue is dialogically relevant. Such dialogical relevance is important to enable new information to be recorded for reuse in future dialogues.

Using this approach the amount of new, typed user input is minimised by allowing the user to select from previously recorded statements first, then allowing the user to type in new statements only if there is nothing appropriate already recorded. The benefit of this approach, as well as maintaining user interest by minimising typing, is that existing statements are reused, possibly in new ways so connections can be made between different threads of argument on a topic. Additionally, this approach avoids the need for natural language processing as propositional statements are recorded in their entirety. When statements are reused in new ways it is because the user has linked the statement to some point expressed within a dialogue. Rich, structured knowledge is thus accumulated through a lightweight, naturalistic interaction with the user

6 Example Dialogue

An example MAgtALO dialogue can be seen in figure 4. In this dialogue there is a single user and two MAgtALO agents, one representing John Wadham of Liberty who is very much opposed to the introduction of identity cards, and the other representing the Labour MP Martin Linton, who argues in favour of identity cards. The dialogue begins at turn 1 from a fixed initial topic. In this case the system introduces the topic of the dialogue with the statement “identity cards are a bad idea”. This does not represent the position of any given participant but merely provides the focus for the dialogue. Once the initial topic has been selected, the user is presented with the option to agree, to disagree, or to find out where the other agents stand with respect to it. In the example the user selects to agree and is then invited to support their position with a reason. In turn 2 the user does so and selects to reuse an existing reason from the system to support their position. Automatic interjection is used to help the dialogue seem more natural and this is demonstrated in turn 3 where the John Wadham agent interjects to agree with the user and proffers a reason for so-doing. This indicate that the John Wadham agent had a strong desire-to-speak in agreement with the user’s last statement.

Once an agent has interjected, the dialogue game allows the user to either agree or disagree with the current point, the last point that was made during the interjection, or to question the agent that made the point to explore that agent’s position. This can be as simple as asking, *Why?*, in order to get underlying reasons and so expose the basis for the agent’s position. If the agent’s point failed to persuade the user, further justification can be solicited. The user then responds in turn 4 by using a dialogue system move to attempt to elicit a further reason from the John Wadham agent who responds in turn 5 with a further reason. It is at this point that the Martin Linton agent automatically interjects to disagree in turn 6 with the statement made in turn 5. The focus of a dialogue generally follows the last point that was made, but by asking for further reasons the user is switching focus back to an earlier point to get extra, independent support for the point. This process of focus switching allows the user to return to any earlier point in the dialogue, simply by selecting the new focus-point from the dialogue transcript displayed on screen. Such a switch of focus is illustrated in turn 7 whereby the user indicates that they wish to return to an earlier point. In this case the earlier point is the initial point of the dialogue and the user further indicates that they wish to explore Marting Linton’s positions. The result of such focus switching is that the user is able to explore new threads of reasoning and expose different arguments for and against each point made rather

than being locked into a particular path through the dialogue. Again this is an example of how the protocol enables a natural rhythm to be maintained in which, when the user is dissatisfied with the current position, they can return to the point of contention and explore it some more. The user, in turn 8, then directs a question towards the Martin Linton agent who responds in turn 9. In turn 10 the users disagrees with the Martin Linton agent but the reason that the user produces is not already available in the system and is a new statement entered into the system through the knowledge elicitation process discussed in section 3. The example therefore demonstrates the major facilities of the MAgtALO system, including knowledge elicitation, automatic interjection, and guidance of the dialogue through the use of a dialogue game protocol.

Turn	User	Statement
1	System	New discussion started! With the point "Identity cards are a bad idea"
2	User	"I agree. They will undermine civil liberties"
3	John	"Yes, I agree. The real target is the people whom the home office has lost"
4	User	"Why do you think the current point is true John?"
5	John	"It will be black people and those from ethnic minorities who will find that they need to show their card most often"
6	Martin	"No, I don't agree. If you are saying the police might discriminate against black people (even though it is illegal), that is a serious problem that needs to be tackled. But should not stop us having ID cards"
7	User	"I'd like to go back to the point that was made earlier...Identity cards are a bad idea"
8	User	"Do you agree with the current point Martin?"
9	Martin	"No, I don't agree. They will not undermine civil liberties"
10	User	"No, I don't agree. Storing potentially vast amounts of sensitive data about every citizen in one place presents dangers"

Figure 4. An example dialogue on the topic of identity cards.

7 FUTURE DIRECTIONS

It has been suggested that argument provides a more intuitive and accessible means of presenting and assimilating complex data [2], and that structured argumentation can be applied to discussions of complex domains involving real risks [5]. In Magtalo, both monologic argument structures and dialogic argument protocols are used to give the user intuitive control over navigation of a complex disagreement space. Presenting and organising material explicitly as arguments should mean that users find it easier to understand the relations between the various positions in comparison to sources which have a more discursive style (such as newspaper reports). One would expect the same to be true for other argument-based systems such as debatapedia. But providing an intuitive interaction metaphor with which the user is expected to be familiar (chairing a meeting), and allowing the user active participation in both directing the discussion and contributing to it, it is further expected that Magtalo should offer an appreciable benefit over formats that allow little or no active participation with the material (such as reports from the traditional media) or that offer a weak, non-argumentative interaction model (such as or wiki pages and discussion boards). Although informal, small-scale evaluations conducted at Dundee suggest that this benefit is substantial, larger scale investigations are required. Testing these hypotheses on specific user groups is a key step for guiding both the Magtalo project specifically, and the online argumentation research area in general.

From a technical perspective there are two key aspects to MAgtALO's future development. The first aspect is to allow the system to use a variety of dialogue protocols. This enables various protocols both to be explored and evaluated using real-world data and for protocols to be developed and deployed that are specific to the needs of the audience. The second aspect is to replace the existing argument processing machinery, currently based upon Araucaria's AML [6], with the ability to import from, and export to, the argument interchange format (AIF) [1] a nascent format for argument representation and interchange. The adoption of AIF enables MAgtALO to become just one system in a possible constellation of online argumentation systems, enabling users to both interact with existing argument resources and to create new ones.

The aim of this research is to exploit advances in argumentation theory by applying them in tools and interfaces that have wide populist appeal. MAgtALO is the first implemented example of an online system that uses a closely specified argument-based dialogue protocol combined with a rich monologic argument representation language to enable users to intuitively explore a *space of diagreement*.

ACKNOWLEDGEMENTS

The implementation work on Magtalo was carried out by John Lawrence, an MSc student in the School of Computing during 2006. John can be contacted at *mail@johnlawrence.net* A video showing how Magtalo can be used, and a live demonstration version of the system, are available online at *arg.computing.dundee.ac.uk*

REFERENCES

[1] C. Chesnevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, and S. Willmott, 'Towards an argument interchange format', *Knowledge Engineering Review*, **21**(4), 293–316, (2006).
[2] John Fox and Subrata Das, 'A unified frameowrk for hypothetical and practical reasoning (2): Lessons from medical applications', in *Practical Reasoning: Proceedings of the International Conference on Formal and Applied Practial Reasoning (FAPR-96)*, eds., Dov M. Gabbay and Hans Jurgen Ohlbach, LNAI 1085. Springer, (1996).
[3] Calico Jack Ltd. http://www.calicojack.co.uk, 2008.
[4] J. D. Mackenzie, 'Question begging in non-cumulative systems', *Journal of Philosophical Logic*, **8**, 117–133, (1979).
[5] P. McBurney and S. Parsons, 'Risk agoras: Using dialectical argumentation to debate risk', *Risk Management*, **2**(2), 17–27, (2000).
[6] C. Reed and G.W.A. Rowe, 'Araucaria: Software for argument analysis, diagramming and representation', *International Journal of AI Tools*, **14**(3-4), 961–980, (2004).
[7] J. R. Searle, *Speech Acts*, Cambridge University Press, 1969.
[8] Douglas N. Walton and Erik C. W. Krabbe, *Commitment in Dialogue*, SUNY Press, 1995.

When the experiment is over: Deploying an incentive system to all the users

Rosta Farzan
School of Information Science
University of Pittsburgh
210 S. Bouquet St.
Pittsburgh, PA 15260
rosta@cs.pitt.edu

Joan M. DiMicco, David R. Millen, Beth Brownholtz, Werner Geyer, Casey Dugan
IBM Research
1 Rogers Street, Cambridge, MA 02142
{joan.dimicco; david_r_millen; beth_brownholtz; werner.geyer; cadugan} @us.ibm.com

Abstract. Motivating user participation is an important issue for the survival of social web and social software applications. In our previous work demonstrated that a point-based incentive encourages contribution to a social networking site. This paper presents a follow-up analysis after a full deployment of the incentive mechanism to the entire user community. We address an issue uncovered in our previous experiment, we measure the long-term impact of the incentive mechanism on site content generation, and we replicate our previous result with a larger number of users. Our results will demonstrate that the incentive mechanism had a long-term effect on contribution levels and generated a second boost in contribution levels when released to a new set of users. The paper concludes with a discussion of our community's reactions to the incentive mechanism collected through the site itself, company-internal blogs, podcasts, and forums.

1 INTRODUCTION

Growth of the social web and of the popularity of social software means that, increasingly, the success of websites and software applications is dependent on user contributions. This raises the important issue of how to persuade users to participate. Prior work, such as our own [6], found through controlled experiments that different incentive mechanisms work on varied conditions. In this paper, we present a follow-up to our previous experiment that presents findings on how a full deployment of an incentive system impacts an on-line community involving thousands of users within the workplace setting.

Our previous research found that, when a points-based system was put into place for half of the users of our social networking web site, those users that received points increased their contributions to the site. There was no corresponding increase in contributions for the users who did not receive points.

After this experiment, there were several remaining questions. First, in our original analysis we observed that many users did not navigate to their personal points page, indicating they did not notice the presence of system, perhaps because there was no explicit announcement of the new site feature. In this research we are looking at the effect of personalized email notification in addition to the points system.

Second, while we found that introducing a points incentive to our original community increased the contributions of the group, we wanted to know if we could replicate that jump in contributions several months later, when the site had several hundred more users, by releasing to the points to the other half of the site's users. We also wanted to know if just before releasing the points we could see a substantial difference between the group who had points for several months and the group that was about to have points shown to them.

Our third and final research question was to ask how is a points-based incentive system received by a community of over 4000 users, as the site measured at the end of our analysis. After all of our site's users could see the points system, we observed that members of the community responded explicitly by gaming the system, discussing and sharing their thoughts on the system through the communication mechanisms in the site, and commenting on it outside of the system, within company-internal blogs, podcasts and forums. The enthusiastic, and sometimes heated, debates about incentive systems highlight some of the pros and cons of the system deployed. The response also highlights how a community of employees working in the software business respond to such systems. Others' work on incentive mechanisms has been deployed to educational communities and onto the open Internet – our study of a large community of employees and their reactions to this system within the workplace is a unique perspective that can inform the design of other planned deployments of incentive systems.

2 BACKGROUND

The appropriate way to motivate user participation always depends on the task, the application, and the users' characteristics. Researchers have looked at different incentive systems addressing different type of tasks and users. Their techniques can be classified into the following approaches:

By rewards: awarding users for their contribution [1,3,4].

By explaining community benefit: highlighting the importance of users' contributions for the community [2,7,8].

By goal-setting: setting a challenging and short-term goal for the users [2].

By reputation: enhancing users' reputation in the community through their contributions [8].

By providing self-benefit: turning users' participation into an important activity for themselves [5].

In our previous work we deployed a point-based incentive system on a social networking site inside of an enterprise, to combine ideas from incenting by rewards and reputation.

The site we deployed the incentive mechanism to was a social networking site designed for IBM employees to network with each other. On the site, users create profile pages, share photos and lists, and comment on each others' content. Persuading users to contribute content to the site is an important piece of our effort to build and sustain a lively community.

Our incentive system rewarded points to the site's users for contributing content to in the form of profile information, photos, lists, and comments on any of the site's content. The number of points for each user was calculated based on the following formula.

$$p = 100 \times (\textit{if any item in profile}) + 15 \times \textit{number of comments} + 10 \times \textit{number of lists} + 5 \times \textit{number of photos}$$

To extend the points incentive from reward to also reputation, we defined four different status levels based on the number of points as shown in Table 1. (The social networking site is called Beehive, hence the bee status names.)

Table 1. Points-based status levels

Number of points	Status level
< 110	New bee
< 500	Worker bee
< 2000	Busy bee
>= 2000	Super bee

To make users aware of their points and status level, and to enable them to compare themselves to others, we implemented the points system in the following ways on the web site:

- The top 10 point-earners were listed on the home page.
- Each user's points and status level appeared next to their name throughout the site.
- There was a page listing the users in each class ordered by number of points.
- And there was a personalized page for users to compare themselves with other people in their network by number of points and status.

We studied the effect of this point-based incentive system in a controlled study, assigning half of the population to an experimental group which had access to points system and the other half to a control group which didn't see any information about points. The study found that the points system motivated the experimental users to contribute all types of content; however the effect was not sustainable and the higher levels of contribution began to drop after one week [6].

At the conclusion of this study, we had three main questions. The first question was what if the users were told more explicitly about the points system? Would a personalized email notification about points make even more users contribute to the site? The second question was can we replicate these results with a different and larger population of users or was this a special group of enthusiastic early adopters? The third question was more general: what will happen when the whole site has the points system visible? Will there be gaming of the system? (We did not observe gaming in our original experiment.) Will users refer to points in their contributions? The next three sections of this paper attempt to answer these questions.

3 EFFECT OF EMAIL NOTIFICATION

Our previous study found that most of the users in our experimental condition (72%) did not visit any of the points pages during our experiment, indicating that they probably did not notice the addition of points system. If users are unaware of an incentive system, the persuasive impact of the system is quite limited. To address this problem, we designed personal notification emails to be sent users about the points system. The site already sent email updates to users about their social network's activity either daily or weekly, depending on users' preferences. To increase the likelihood that users would discover the points system, we added the following information to the top of these personalized email notifications:

> You are a new bee with 25 points:
>
> You only need 85 more points to become a worker bee.
> - Each time you share a photo, you earn 5 points.
> - Each time you create a hive5, you earn 10 points.
> - Each time you comment on a profile, a photo, or a hive5, you earn 15 points.
> - The first time you put content into the "about you" section of your profile page, you earn 100 points.
>
> Click here to find out the details about the beehive points and compare yourself with your network.

The highlighted portions of the above text were personalized for each user, showing status, total number of points, and number of points needed to jump to next class. The message about adding information into the profile was included if the user had no profile information.

To evaluate the effect of this addition to the email notifications, we looked for changes in contribution levels within the experimental group (41 users), as compared with the control group who did not see any changes to the site or their emails (40 users). Looking at users' total point values, in the week following the email notification, 12% of users in the experimental group moved out of zero points status and 7% of them jumped from New bee to Worker bee status, while only 5% of the control group moved out of zero-point status and no one in the control group earned enough points to jump from New bee to Worker bee.

Figure 1 compares the control versus experimental groups in terms of the average number of photos, lists, and comments added the week before and the week after receiving the first enhanced email notification. As the data shows in Figure 1, within the experimental group, there was a 67% increase in the number of photos added, a 75% increase in lists, and a 92% increase in comments. Figure 2 shows the percentage of users in each group who added content to their profile section for the first time, the week before and the week after the email was sent.

These results suggest that the email including notification about the points caused an increase in the amount of content added by the experimental group. However, the difference is not statistically significant for each content type separately or for the overall amount of points earned by each group (We ran a repeated-measures ANOVA, considering time as the repeated

measure, group as the independent variable and total amount of new content added to the site as the dependent variable).

There is further evidence to support that the email encouraged users to contribute content found in the percentage of users contributing to the site each week. We found that the percentage of users contributing photos and lists nearly doubled in the week following the email notification (7% to 12% of users for photos, 5% to 10% of users for lists); the percentage of users adding comments dropped by 2% from 24% to 22%.

Our conclusion from this analysis is that notifying users by email about the points was an effective way to extend the awareness of the points system and in turn generate more user content.

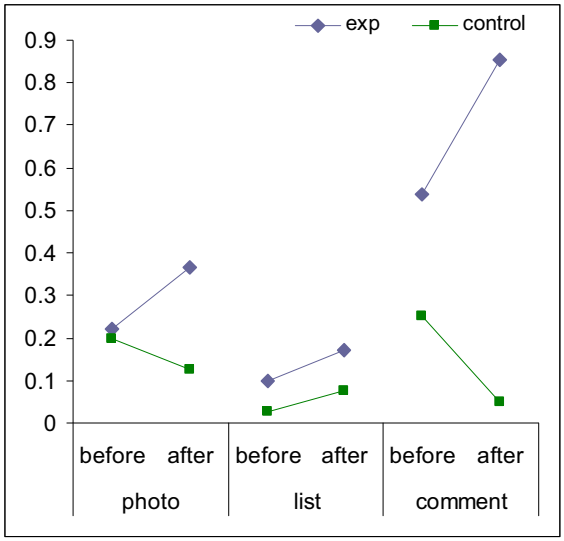

Figure 1. The average number of photos, lists, and comments added by each group the week before and the week after the email notification

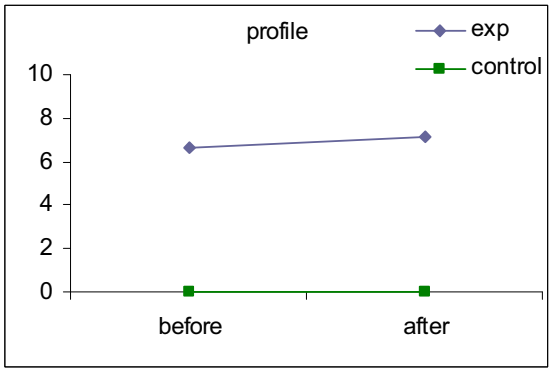

Figure 2. Percentage of users in each group who added content to their profile "about me" section for the first time

4 RELEASING POINTS SYSTEM TO ALL USERS

At the conclusion of our experiment, we left the points system visible to just half of the users for another three months. At that point, after the site's community had more than doubled in size, we determined it was time to let all users see the points system, particularly since the points and the email had generated boosts in site contributions.

By releasing the points to the entire site, we can also assess two aspects of the points system. First, we can asses the long-term effect of one group having points and another not. Second, we can see if our original findings, that points generated a boost in contributions, are repeatable, with a new, larger community, when we release the points to the control group.

In our previous experiment [6], the experimental group had 63 users, as did the control group. Now the group that had the points had grown to 207 users and the group that would be newly introduced the points had 214 users. For clarity, we will refer to these two groups in our analysis as **PtsGroup1**, the group of users who had access to the points system for several months and in most cases since the day they joined the site, and **PtsGroup2**, the group who the points were released to all on a single day.

Our two evaluations were done through log analysis considering six weeks of usage logs – the three weeks before and the three weeks after releasing points system to PtsGroup2. For consistency we limited the users of the study to those who used the system consistently over the six-week period, which means they used the system at least once every week.

4.1 Differences between PtsGroup1 and PtsGroup2

Before releasing the points, we assessed the long-term impact of the points by looking for differences between PtsGroup1 and PtsGroup2. Table 2 presents the amount of content added by the users in each of these groups, both before releasing the points and three weeks after the points were released. From the numbers, we can see that PtsGroup1 added much more content to the site than PtsGroup2, especially in the case of comments, where users added three times the number of comments to the site. Fitting negative binomial regression model shows a significant effect of group for lists and comments (lists: df=418, 1, χ^2 =6.38, p=0.01, comments: df=418,1, χ^2 =25.14, p<0.0001). These data provide strong evidence that the points system generates more content on the site over the long term.

Table 2. Differences in contribution amounts between PtsGroup1 (207 users) and PtsGroup2 (214 users)

	Total Photos	**Total Lists**	**Total Comments**
Before releasing points:			
PtsGroup1	823 (48% more)	443 (74% more)	2703 (299% more)
PtsGroup2	556 (32% fewer)	254 (43% fewer)	678 (75% fewer)
Three weeks after:			
PtsGroup1	991 (37% more)	484 (49% more)	3630 (234% more)
PtsGroup2	722 (27% fewer)	324 (33% fewer)	1086 (70% fewer)

Table 3. The percentage of users in each group that had contributed content before the release of the points to PtsGroup2

	Photos	Lists	Comments	Profile "About Me"
PtsGroup1	44.4%	43.0%	42.1%	43.9%
PtsGroup2	45.6%	38.3%	43.7%	36.4%

Table 3 presents the percentage of users who contributed content to the different parts of the site. These percentages are all roughly the same: between 36 and 46% of users are contributing photos, lists, comments, and profile information to the site. These data indicate that the points system over the long term did not encourage more users to contribute to the site. So at this point in our analysis of the system, it appears that the long term impact of the points system is that users contribute *more content* to the site but the same number of users contribute as would without the points system.

4.2 Impact of releasing points

In releasing the points to the entire site so that PtsGroup2 would see the system, we hypothesized that we would observe a jump in the amount of content added by PtsGroup2, replicating result observed in our previous study. Comparing the contribution of the groups the week before and after releasing the points finds significant interaction between time and group in terms of number of points earned (df=1, 418, F=4.19, p=0.02), number of added lists (df=1, 418, F=5.35, p=0.01), and number of added comments (df=1,418, F=3.71, 0.02), but not significant for photos (df=1,418, F=0, p=.98) (These tests were repeated-measures ANOVAs, considering time as the repeated measure, group as the independent variable and total amount of new content added to the site as the dependent variable). The results, in Figure 3, show that the group new to points system started adding more content, earning more points. These results charts are very similar to those from our previous experiment, so it appears that the release of the points had the same effect as the introduction of the points to the initial, smaller group.

Focusing on just the content added by PtsGroup2, we compared different content types added by them over time. As mentioned before, users received the highest number of points for adding comments, followed by lists. As a result, we expected to observe a more significant increase of contribution in the form of comments, and lists. The result is shown in Figure 4, supporting our expectation that the rate of growth of comments and lists are higher than photos.

Our final analysis of the effect of releasing points to PtsGroup2 was looking at the percentage of users in each group that contributed the different content types. As shown previously in Table 3, there was little difference between the two groups before the release of the points: between 36% and 46% of users were contributing in both groups to all categories of content. This may indicate that the points system encourages quantity of contributions, rather than encouraging new users to contribute to the site. But there is now evidence to dispute that, at least in the short term. Three weeks after releasing the points to PtsGroup2, the number of users that added new content to each content type is much higher than in PtsGroup1. Between 3.4% and 6.4% of the PtsGroup2 users added content to site for the first time, compared to 0.9 to 1.9% of PtsGroup1. This indicates that in the short term, the points system encouraged users to add content for the first time. It may be that this finding reveals a distinction between short-term and long-term effects of incentive systems: that short-term they encourage new users to contribute, but over the long-term those users would have contributed anyway. Our consistent finding is that this incentive system encourages users to share more content than they would have otherwise.

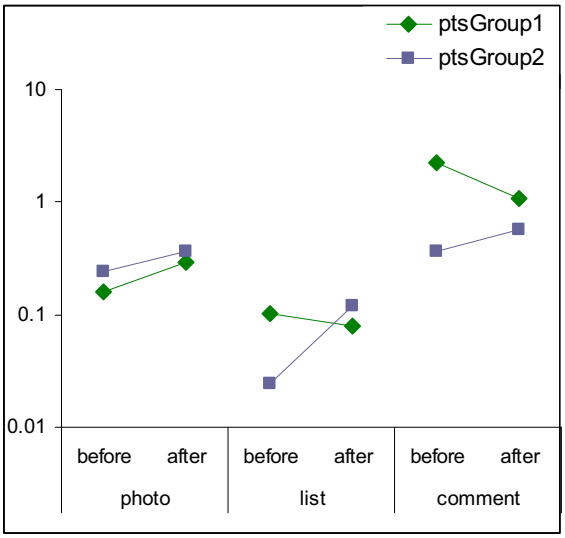

Figure 3[1]**.** The average number of photos, lists, and comments added by each group each week

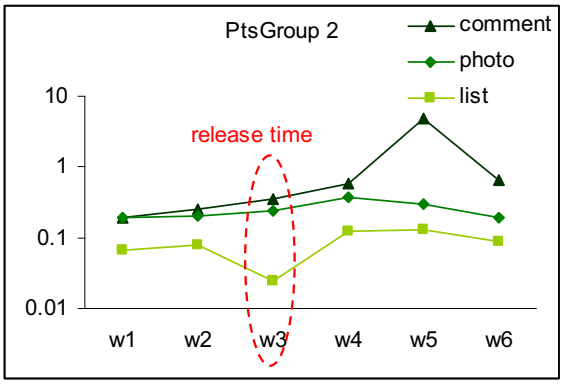

Figure 4[1]**.** Comparing the amount of each content type added by PtsGroup2 over time

Table 4. During the three weeks after releasing points, the percentage of users in each group to add photos, lists, and comments *for the first time*.

	Photos	Lists	Comments	Profile
PtsGroup1	0.9%	0.9%	1.9%	1.4%
PtsGroup2	3.4%	6.4%	5.3%	5.3%

[1] For better presentation, the graphs in Figure 3 and 4 are shown on a logarithmic scale.

5 USER REACTIONS TO THE POINTS

Shortly after the points system was released to the entire community, members of the community began to mention the points and the different bee levels on their profiles, in the comments, in their status messages, and as the topic of discussion for lists. While most of the site content relating to points can be summarized as users talking about their own point level and bee status level, a vocal minority began discussing the points system in general, having a lengthy discussion of the pros and cons of the system and suggesting alternatives. Another small set of users openly gamed the system by adding low-value content to the site for the purpose of gaining points. This section describes these different explicit responses to the system. When we performed this analysis, the points system had been released to the entire site for two months and the system had over 4000 users.

5.1 Discussing merits of the points system

The most visible discussion of the merits of the points system occurred within seven different lists on the site that had 87 comments made on them. Outside of the site, there was also a blog post with 19 comments, a discussion forum thread with 20 posts that both debated the pros and cons of the system, and a company-internal "social software podcast" invited on one of the vocal members of these discussions specifically to lead a conversation during the podcast about the drawbacks of our site's points system. All of these discussions involved 39 people, not counting the site's project team members. (We, the team, posted minimally to the discussion, only to clarify the mechanics of the system and ask for clarifying points from the users on what they would like to see and to understand which parts of the system were particularly objectionable to users.)

The main arguments levelled against the points system were:

Users cannot opt-out. This point was raised by some of the highest point earners involved in the discussion. They felt they should have a choice as to whether or not their activity level was revealed on the site. They felt the visibility of their activity level was a violation of their privacy. Others stated that they personally would not opt-out of the system, but they agreed with this opinion and wanted this ability added to the points system.

Points do not reflect quality or meaning of the contributions. There was a lot of discussion about how quantifying contributions lessens the meaning of the contributions. An ideal system would measure the quality of the contributions in more qualitative ways, which users admitted was difficult to do. As an alternative opinion, some felt that any incentive system detracted from the real "value" of the site, which is the personal connections made between people. Those who agreed with this would prefer the site have no incentive mechanism in place.

Desire for customizable systems and points for more actions. A small number of users suggested that users be able to craft the incentive systems by assigning point values to the things they felt were most valuable. Others wished users could earn points for "friending" because that was what they perceived as the most valuable activity.

The users involved in the discussions who defended the points system also made some important points about its limitations:

When getting started, the points system can be motivating, but after you attain a high level of points, the incentive no longer works. The users who felt this way were high-point earners, contributing to many discussions elsewhere on the site, and connecting to many people. So while they stated that the points were "meaningless" to them, they also stated that they felt when they were starting out on the site, the points were a helpful way to gauge progression through the site.

The points help users distinguish experts from newbies. A couple of users specifically approved of the label "New bee" as a good label to attach to users who were new to the site and perhaps needed encouragement to start contributing. The "Super bee" label was mentioned by others as a way of finding out which users were experienced with the site and heavy users.

Improve the points with a decay function. An issue that had come up in our earlier paper's user interviews [6] and was also brought up in this discussion was that users would like to see the points decay over time. The main reasons for wanting points to decay or reflect time in some way was that they wanted to have a way to detect recent high contributors to the site, rather than the current model which highlights the highest contributors over a long period of time.

From these valuable discussions, we learned that our users have well thought-out opinions about incentive systems and their role within an enterprise environment. They have contributed to our overall understanding of these systems. We also agree with many of their suggestions and plan to implement the opt-out feature and possibly a time-based decay function on the points.

5.2 Gaming the points system

Most of the discussion on the site about the points systems was led by a single user who felt passionately that users should be able to opt out of the points system. To demonstrate his position that the points system should not be a representation of a person's contribution to the site, he launched a keyboard macro that submitted comments onto one of his photos with the single word "test." This generated 1000's of points for his profile and raised him to the second highest point position on the system. (He stopped the macro just short of the top position because he was friends with the person in the top position and said he did not want to offend him.) Once the flurry of activity around the points discussion died down, this user deleted the photo that had all of these "test" comments, and thus returned his point value back down to a lower level.

In addition to this example of gaming, we have evidence of three other instances of gaming the points system.

Two users earned 100 points each by adding content to their profile that mentioned the points system and did not provide any information about themselves. These two profile self descriptions were "What do you think about putting this here to get the 100 points I needed to be a nicer bee than a noobie?" and "Why [am I] entering a question here? I read on the newsletter that you will get 100 points and hence here I am...all part of the master plan to becoming a Queen Bee)."

The third instance of gaming was a user who posted the exact same comment on 21 different profiles. The comment said "You have zero beehive points. Maybe that's something to work on, so you don't look like such a newbie. Thanks for letting me make points on you just for posting this comment!"

Overall, these instances of gaming seem fairly benign considering there were over 4000 users on the site at the time of our analysis. In designing any incentive mechanism, it is always a challenge to design one that discourages gaming. Although the points system is technically easy to game, we believe it is *socially* difficult to game because users are communicating with their colleagues every time they post content. By gaming the system, one's coworkers become acutely aware of the gaming because of the system's network notifications.

As was brought up in the discussions about the merits of the points system though, there is an issue of low quality contributions being counted equally with high quality contributions. We did not assess the quality of the comments being left on the site because the quality of communication between two users cannot be judged by an external party. For example, a simple "hi" can convey a lot of meaning between colleagues, which an automated analysis would miss. Therefore we have not been able to assess whether the quality of the comments, photos, or lists left on the site has decreased due to the points system.

5.3 Mentioning one's own points

Outside of the discussion about the merits of the system, there is evidence throughout the site that users were aware of the system and had a desire to earn more points. Sixty-five people commented about their points in a variety of locations on the site: 21 people commented on people's profile pages about their own or their colleague's points; two people mentioned the points in their profile self descriptions (in addition to the two who gamed the points); one person used his list's comment section to test out how points were earned; two people talked about earning points while adding a comment about photos; and 37 people mentioned points in their status messages.

In interesting aspect to the status messages is that one does not earn points for changing one's status, so these comments were not motivated by a desire to earn points. Most of the status messages were about one's point level: "just a new bee :(," "a simple worker bee," "figuring out how to get points," "just over 800 (make that 900) points, narf, narf," "deciding whether to be a busy bee or not," and "wow, Ricardo has 15000+ points, and I have 80... suddenly, I feel inadequate..."

5.4 Summary of employees' responses

At the conclusion of this several week interaction within the community, we believe there are several important lessons to be learned about introducing incentive systems to an existing site. First, established community members, given the opportunity to give feedback, will have strong opinions on the introduction of a feature such as this that ranks users and establishes an evaluation criterion. Second, much of the feedback provided by the community was insightful and the suggestions for changing the system were provided from the perspective of users within a larger ecosystem, our enterprise setting. So we believe that the suggestions made for changing the points system should be taken very seriously. Based on their feedback, we plan to provide an opt-out ability so that users can elect to not have their point levels shown on the site. Our third lesson is that a small feature does have an impact on how people communicate on the site, from the gaming to the joking references about the points. Discussion of the single feature was pervasive and was not limited to the features of the site that granted points.

6 CONCLUSION

Our qualitative and quantitative results show that overall the points system is a successful method for motivating users' contributions in an enterprise social networking: it boosts users' contributions when introduced, particularly when paired with an email notification, and over time the users who have points visible to them contribute much more content to the site, albeit not at higher levels of user participation. Furthermore, it also stimulates discussion amongst the users.

A broader implication of this finding is that if designers of a site wish to generate a burst in user activity, the introduction of a simple points-based system can accomplish that. Therefore, in addition to increasing the content on the site overall, the launching of a points system can be seen as a way of controlling the behavior of the users to generate a predictable result at a specific moment in time. This gives an opportunity to site owner's to target users and could have value for marketing efforts.

7 REFERENCES

[1] Bretzke H., Vassileva J. (2003) Motivating Cooperation in Peer to Peer Networks, In Proc. Of UM03, Springer Verlag LNCS 2702, (2003), 218-227.

[2] Beenen G., Ling K., Wang X., Chang K., Frankowski D., Resnick P., Kraut R.E., Using social psychology to motivate contributions to online communities, In Proc. CSCW 2004, ACM Press (2004), 212-221.

[3] Cheng R., Vassileva J. Design and Evaluation of an Adaptive Incentive Mechanism for Sustained Educational Online Communities. User Modeling and User-Adapted Interaction, special issue on User Modeling Supporting Collaboration and Online Communities (2006), 16 (2/3), 321-348.

[4] Cheng, R., Vassileva, J.: User motivation and persuasion strategy for peer-to-peer communities. In Proc. of the 38th Hawaii International Conference on System Sciences (Mini-track "Online Communities in the Digital Economy"), IEEE Press (2005), 193–202

[5] Farzan, R., and Brusilovsky, P. Social navigation support in a course recommendation system. *In: V. Wade, H. Ashman and B. Smyth (eds.) In Proc. AH'2006, Springer Verlag (2006)*, pp. 91-100

[6] Farzan, R., DiMicco, J. M., Millen, D. R., Brownholtz, B., Geyer, W., and Dugan, C. (2008). Results from Deploying a Participation Incentive Mechanism within the Enterprise. To be appear in proceeding of CHI2008.

[7] Rashid, A. M., Ling, K., Tassone, R. D., Resnick, P., Kraut, R., and Riedl, J. Motivating participation by displaying the value of contribution. In Proc. of CHI 2006, ACM Press (2006). 955-958.

[8] Kollock P., The economies of online cooperation: gifts and public goods in cyberspace, In: Smith, M.A., Kollock, P. (Eds.), Communities in Cyberspace. Routledge, (1999) London.

[9] Ling K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frakowski, D., Terveen, L., Rashid, A. M., Resnick, P., and Kraut, R. Using social psychology to motivate contributions to online communities. JCMC, 10(4), 2005.

Encouraging Community Spirit with Situated Displays

Nick Taylor and **Mark Rouncefield** and **Keith Cheverst**[1] and **Shahram Izadi**[2]

Abstract. There is a widespread belief that 'community' has been greatly diminished in many areas, partly attributed to the isolation and weak social ties encouraged by modern technology such as television, computer games and the Internet. We explore a photo display application which may help to reinforcing the community's values, support community ties and integrate individuals into the community.

1 INTRODUCTION

Communities, small-scale social groupings of various kinds, appear crucial to social life. Consequently, arguments about the impact of social, economic and technological change on the nature, importance and influence of these close-knit communities—the 'decline' of community—have engaged commentators for some time [15, 18]. Often, modern technology is seen as being at least partly to blame; widespread car ownership has broken down the geographical boundaries around communities and the reliance on nearby resources, while telecommunications enable relationships to be formed regardless of physical location.

Although the broadening of social life could easily be seen as a positive change, it has been argued that the long distance, anonymous relationships—or 'weak' ties—afforded by the Internet are superficial and a poor substitute for the 'strong' ties provided by local communities and social networks [4]. These close relationships are believed to be an important part of our lives; social disengagement has been associated with crime, inefficient government and poor physical and mental health [12]. This does not mean that weak ties are inherently flawed or without merit. Studies have shown that they may be equally important in communities, allowing cross-communication and information sharing between tight-knit groups [11], strengthening the community as a whole. For example, it has been argued that these links may play a key part in enabling grass-roots action for common causes [6].

The ties afforded by networked technologies have also proved to be particularly adept at maintaining strong ties where they already exist [19]. Previous work in this direction has already demonstrated the potential for technology in supporting community, particularly through 'community networks'—electronic social networks which have grown from an existing geographic community [16]. One of the better known examples of this kind of system, the Blacksburg Electronic Village [2], led to reports from community members of increased levels of communication and greater participation in and access to community activities. Likewise, studies of the networked community in Netville showed an increase in recognition of neighbours by those connected to the network after the same length of residence [8].

By allowing public access to electronic resources in-situ, situated digital displays also show promise in promoting community. For example, the Community Wall system [7] aimed to improve workplace communities by displaying content of common interest to stimulate conversation between employees who might not normally speak to each other. The eyeCanvas system [3] has also proved to be a popular installation within an art community's social space, allowing playful notes and in-situ access to artists' work.

'Community' is, of course, a complex social construct, but without over-hyping the technology, we believe that public digital displays, designed with the community in mind and deployed in key social spaces, may well be able to help support communities by fostering notions of community identity and shared history. To investigate this, we have developed the Wray Photo Display [17], a public photo gallery application deployed in a small rural village.

With this system, we hope to demonstrate how a digital display of photographs may be able to support communities through reinforcement of community values and by demonstrating the activities, history and other parts of the community's tapestry which are most important to its members. The user-generated nature of this content could further allow individuals to impart their own views on exactly what those community values are. This might serve not just to strengthen the commitment of existing members, but also to advertise the benefits of community to those who currently fall outside its boundaries, including visitors and perhaps most importantly, to new residents.

Figure 1. The Wray Photo Display.

This paper explores the Photo Display as a potential application for persuasive systems and presents our initial findings on its use and effects. In this we are attempting to move away from crude characterisations of community or even cruder understandings of the impact of technology—instead we are trying to point to and develop

[1] Lancaster University, UK. Email: n.taylor@comp.lancs.ac.uk, m.rouncefield@lancaster.ac.uk, kc@comp.lancs.ac.uk
[2] Microsoft Research, Cambridge, UK. Email: shahrami@microsoft.com

more subtle understandings both of what a 'community' might be and what a 'persuasive' technology might achieve in this context. Thus we attempt to engage with some of the subtleties required for understanding 'persuasion' as a process designed to involve the user, to inveigle, to entrance, to charm, and what part technology might reasonably play in this process.

2 THE WRAY PHOTO DISPLAY

Our studies into community have concentrated on the village of Wray (Figure 2), a rural community 15km north-east of Lancaster with a population under 500. The village has comfortably integrated technology into everyday life; since 2004, it has been home to a wireless mesh network installed as part of another university project. Both the goodwill generated by this project and the high connectivity available in the village have made Wray an ideal test site for community systems.

After a period of observation to determine potential uses for displays in the village, we opted to use a technology probe approach [10], developing a simple prototype application to explore the deployment environment, field test the technology and elicit ideas and requirements from residents. During the development of the system, we have employed an agile, iterative design model in which new features are developed rapidly as they are requested and deployed into the wild for evaluation.

Figure 2. Wray, Lancashire.

The Wray Photo Display was deployed in August 2006, consisting of a touch screen display (Figure 1) which scrolls through pages of photograph thumbnails uploaded to an associated website by residents. Touching a thumbnail opens a full-screen image and users can browse through individual photos or between photo categories, which residents can create and opt to moderate. Recent developments have added the ability to upload short video clips, browse though images using the website and add comments using an on-screen keyboard.

Initially, this was deployed in the village hall, which was suggested during the earliest design sessions and seen as a central social space in the village community. This building provided a variety of functions, including a visiting doctor's surgery (including a waiting room area) and a local cinema, although the hall was only open to the public while in use. Due to electrical work in the hall, the display was later moved to the village post office, as seen in Figure 1, where it gained considerably more exposure.

Throughout the deployment, we have attempted to be as inclusive as possible—in particular, it remains important to avoid a 'rich get richer' effect, where those most involved with the community benefit most. The public placement of the display (as opposed to an entirely web-based system) ensures that it will be seen by a greater number of residents, including those without access to a PC or Internet connection and we are currently considering solutions for those without computers who wish to post photos.

3 OBSERVATIONS

Our data from the deployment has been received through several channels. Firstly, the display's contents and usage have been monitored remotely, offering a revealing insight into the community and an indicator of its uptake. We have also gathered feedback throughout the deployment from a paper comments book left with the display, as well as design workshops and discussion with members of the community. Finally, we have been able to directly observe users interacting with a duplicate display at popular village events, such as the annual Wray Fair.

3.1 Display Usage

The Photo Display has been well accepted into the community and has seen widespread use, particularly following its move to the post office. To date, just over 800 images have been uploaded, covering village history, scarecrow festivals, day trips, children's artwork and a charity maggot race, and logs show around 300–500 image views a month through the situated display (Figure 3). Usage was particularly high in October and November 2006 following the display's relocation to the post office and increased visibility.

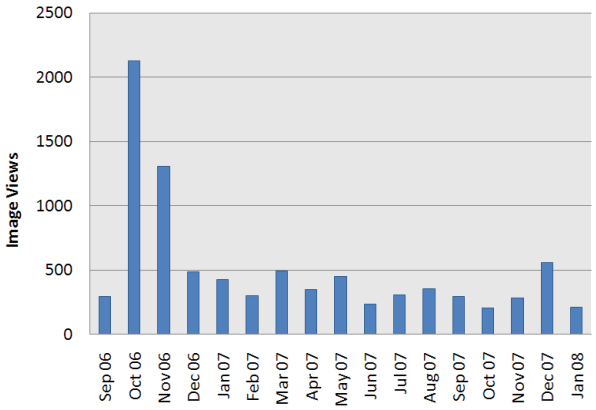

Figure 3. Image views per month.

The residents' choice of categories for images was also revealing. With the exception of a small number of photos from our previous visits to the village, uploaded to provide some initial content, we made no attempt to dictate what content should be posted on the display. With the exception of just 37 images in a 'Funny Pictures' category, all categories and images were related to community events and history.

The 'Old Photos' category, containing historical images dating from as far back as the 19th Century and more recent images depicting current residents as children, is by far the most populous category, containing 188 images. Other popular categories include those for the 'Scarecrow Festival 2006' (126 images) and '2007 Produce Show' (125 images). Many smaller categories grouped under

'Current Events' totalled 164 images, showing various recent community events. It should be noted that the Scarecrow Festival, the highlight of Wray's social calendar, demanded its own top-level category, showing the importance of this event to community members, and providing even greater visibility to outsiders and newcomers.

Comments left on images have largely tended towards identifying people shown in the images, particularly on historical photos and group pictures. The majority of these have been left using the situated display's on-screen keyboard rather than through the web interface.

3.2 Feedback

Feedback has been highly positive and often replete with functionality suggestions which fall both within and far outside the scope of the system, providing inspiration for improvements to the Photo Display and plans for future display deployments in the community. Many of the comments left have had an effect on the current implementation—categorisation, browsing via the website and comments have all been provided in response to feedback.

Over 70 comment book entries have been left so far; many noted that the display was a useful source of information about current events, particularly for new residents and visitors:

> "What a superb idea, especially for those who are new to the village."

> "A lovely way to see what's going on for locals and visitors."

This was also seen as an advantage for members of the community who had been absent, allowing them to still 'participate':

> "I missed the last couple of days of the Scarecrow Fest and this gives me the opportunity to see some of the activities and scarecrows I missed."

Several early comments requested historical photos be posted to the display, or suggested other possible features, many of which were community related:

> "Would be great to see some of the historical pictures of the village and forward notice of village events. What about selling advertising space to villagers, proceeds to a village charity?"

One additional comment received by email very effectively summarised the feedback we've received so far:

> "The digital notice board has many advantages for the village... there are quite a few new people in the village and this gives them an insight as to what Wray used to look like, although visually it has not changed very much. The flood photos are one way the old and newer village can be seen. Also the photos of the previous villagers i.e. school photos, weddings, industries carried out in Wray (which many newcomers will probably not be aware of) and just local characters are invaluable in the history of Wray."

3.3 Direct Observation

Observation of community events has also been invaluable, allowing us to meet with regular contributors and gain an insight into community life. It has served to demonstrate how seamlessly the display can play a part in everyday life; during one community event, a new resident in the village spent some time browsing historical photos, hoping to find an image of his house (Figure 4). He was introduced to a local history enthusiast who helped identify the building and its past uses and inspire a potential name for the house—the presence of the display had facilitated a discussion on community history between two strangers and potentially helped to bring an element of that history back into the present.

Of course, the public nature of the display also lends itself to controversy. During the 2007 scarecrow festival one scarecrow was deemed offensive by several residents and removed—however, photographs had already been taken and were later posted to the display, causing a minor furore.

Figure 4. Two residents discuss local history around the display.

4 DISCUSSION

'Community' has little to do with the individual's geographical location, but is an achieved social construct, a 'persuasion', of mutual ties, orientations and obligations, pointing to the ability of technology to reshape and redefine how people see themselves [14]. What appear to be important features of communal life concerning boundaries and membership, rhythms and relationships, temporality and change, interaction styles and preferences are not naturally occurring features but objects of persuasion, of convincing people to behave, to belong, to relate. Part of the persuasiveness of the technology refers to that nuanced understanding of place and its relationship to community and social practices: the'situated' notion of 'place' that Harrison and Dourish define as "a space which is invested with understandings of behavioural appropriateness, cultural expectations, and so forth." [9]

This complexity means a sense of community is a difficult phenomenon to quantify, given its dependence on a vast number of everyday variables and even the difficulty in reaching a common perception of what constitutes community. However, we believe our observations demonstrate that the Photo Display has a clear potential for reinforcing values, encouraging participation and building a shared history. While we do not envision this kind of display causing major reversals of opinion, we do see a more subtle form of persuasion in action, making a small but significant difference to the way people see their community and see themselves within it.

Wray may already have a strong existing community spirit and high levels of involvement, but feedback and observations of the Photo Display application have shown various benefits to the community, such as examples of the display encouraging interaction between newcomers and established community members. In another

case, historical photos which were once kept in private collections are now available to the public and are used by local school children for research, helping to pass on the community's history to a new generation.

Many of the comments received seem to evoke community features identified by Mynatt *et al.* [14]: comments referring to the integration of new residents suggest membership and apprenticeship, relating to the notion of boundaries, while the popularity of historical photos strongly supports the notion of change and community history—several residents have commented that the display is a "living history" of the village. We have also seen that the user-generated content added to the display offers insights into the community itself, identifying the events and pieces of history that the community sees as important.

Above all, the turn to user-generated content highlights the way in which a sense of community is accomplished and achieved 'in the doing', by putting up photos of village activities and thereby actively reminding a community of their history and mutual ties and obligations etc. Our research suggests that the situated display can operate as elements of Fogg's functional triad [5]: as a tool, as a medium and as a social actor performing subtle forms of persuasion through influencing people's attitudes and behaviour and by providing information. However, there is no 'simple' persuasion here, no propagandising technology and a gullible public; instead people have to be charmed, to be inveigled, to be intrigued, to be persuaded into communal ways of living and the boundaries, relationships and changes this necessarily requires.

The deployment also illustrates some of the complexities involved in getting a 'persuasive' technology to work—that for a technology to be persuasive, people have to be persuaded to use it and that 'becoming a user' involves a myriad of both social and technical subtleties that go beyond simple interface design to an appreciation of what it means to embrace the use of a technology. Lie and Sorenson stated that "when studying technologies we are looking for types of use, symbolic expressions and personal attachment remaking the technologies into something close and familiar. This is a way of making them part of everyday life, and it is not accomplished simply by letting them into the home or other daily surroundings." [13]

In *Community: Seeking Safety in an Insecure World*, Bauman writes that "community is nowadays another name for paradise lost—but one to which we dearly hope to return, and so we feverishly seek the roads that may bring us there" [1]. While clearly technology alone is no easy way back to the 'paradise lost' of community, this study reveals some of the myriad technologically mediated ways in which, notions of community, of communal history, of membership, of belonging and responsibility are continuously asserted and reinforced. Nevertheless, whilst still wishing to avoid the hype often associated with new technology, it is our belief that the affordances of the technology in this particular deployment have a far wider relevance and could be used to strengthen struggling communities, such as deprived urban environments or struggling rural communities, by encouraging participation by residents and thereby actively promoting a sense of community.

ACKNOWLEDGEMENTS

This work is supported by the EPSRC funded CASIDE project (grant ref: EP/C005589) and a Microsoft Research PhD Scholarship. We would also like to thank Chris Conder and the continually helpful and enthusiastic residents of Wray (www.wrayvillage.co.uk).

REFERENCES

[1] Z. Bauman, *Community: Seeking Safety in an Insecure World*, Polity, 2001.

[2] J.M. Carroll and M.B. Rosson, 'Developing the Blacksburg electronic village', *Communications of the ACM*, **39**(12), 69–74, (1996).

[3] E.F. Churchill, L. Nelson, and G. Hsieh, 'Café life in the digital age: augmenting information flow in a café-work-entertainment space', in *Proc. CHI 2006*, pp. 123–128, (2006).

[4] J.N. Cummings, B. Butler, and R. Kraut, 'The quality of online social relationships', *Communications of the ACM*, **45**(7), 103–108, (2002).

[5] B. J. Fogg, 'Persuasive technologies', *Communications of the ACM*, **42**(5), 26–29, (1999).

[6] M. S. Granovetter, 'The strength of weak ties', *The American Journal of Sociology*, **78**(6), 1360–1380, (1973).

[7] A. Grasso, M. Muehlenbrock, F. Roulland, and D. Snowdon, 'Supporting communities of practice with large screen displays', in *Public and Situated Displays: Social and Interactional Aspects of Shared Display Technologies*, eds., K. O'Hara, M. Perry, and E. Churchill, 261–282, Kluwer Academic Publishers, (2003).

[8] K. Hampton and B. Wellman, 'Neighboring in netville: How the internet supports community and social capital in a wired suburb. city and community', *Computer Supported Cooperative Work*, **2**(4), 277–311, (2003).

[9] S. Harrison and P. Dourish, 'Re-place-ing space: the roles of place and space in collaborative systems', in *Proc. CSCW 1996*, pp. 67–76, New York, NY, USA, (1996). ACM.

[10] H. Hutchinson, W. Mackay, B. Westerlund, B.B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderbäck, 'Technology probes: inspiring design for and with families', in *Proc. CHI 2003*, pp. 17–24. ACM, (2003).

[11] A. Kavanaugh, D. D. Reese, J. M. Carroll, and M. B Rosson, 'Weak ties in networked communities', in *Proc. C&T 2003*, pp. 265–286, (2003).

[12] R. Kraut, M. Patterson, V. Lundmark, S. Kiesler, T. Mukophadhyay, and W. Scherlis, 'Internet paradox: A social technology that reduces social involvement and psychological well-being?', *American Psychologist*, **53**(9), 1017–1031, (1998).

[13] M. Lie and K. H. Sorenson, *Making Technology Our Own? Domesticating Technology Into Everyday Life*, Scandinavian University Press, 1992.

[14] E.D. Mynatt, V.L. O'Day, A. Adler, and M. Ito, 'Network communities: Something old, something new, something borrowed…', *Computer Supported Cooperative Work*, **7**(1-2), 123–156, (1998).

[15] R.D. Putnam, *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster, 2000.

[16] D. Schuler, 'Community networks: building a new participatory medium', *Communications of the ACM*, **37**(1), 38–51, (1994).

[17] N. Taylor, K. Cheverst, D. Fitton, N.J.P. Race, M. Rouncefield, and C. Graham, 'Probing communities: Study of a village photo display', in *Proc. OzCHI 2007*, pp. 17–24, (2007).

[18] B. Wellman, 'From little boxes to loosely-bounded networks: The privatization and domestication of community', in *Sociology for the 21st Century: Continuities and Cutting Edges*, ed., J.L. Abu-Lughod, 94–114, University of Chicago Press, (1999).

[19] B. Wellman, A.Q. Haase, J. Witte, and K. Hampton, 'Does the internet increase, decrease, or supplement social capital? social networks, participation, and community commitment', *American Behavioral Scientist*, **45**(3), 436–455, (2001).

A Dominance Model for the Calculation of Decoy Products in Recommendation Environments

A. Felfernig[1], B. Gula[3], G. Leitner[2], M. Maier[3], R. Melcher[2], S. Schippel[1], E. C. Teppan[1]

Abstract. Recommender systems support internet users in finding and identifying products and services suiting their individual demands (e.g.: digital cameras, financial services). Although recommender systems already offer mechanisms which alleviate the comparison of different suitable products (e.g. product lists, comparison pages) users usually have difficulties in decision making and finding the optimal option. Persuasive mechanisms can be used in such situations for underlining product differences and reinforcing confidence in the users' own decision. This leads to an increase of the trust level and supports the decision making process. Especially theories concerning user behaviour in buying situations constitute great potential for persuasion in recommender systems. It has been shown that the user's perception of the value of a certain product is highly influenced by the context (i.e. the set of presented products in the choice set). Well-known context effects are the asymmetric dominance effect, the attraction effect, or the compromise effect. This paper presents a multi-preferential and multi-alternative model (i.e. more than two product attributes and more than two products are supported) for calculating dominance values of items in choice sets and thus offers the possibility of determining the best recommendation set in a given choice situation. The performance of the model is shown by the application on empirical data (choice sets) gained by a previously conducted user study.

1 INTRODUCTION

Due to the complexity and size of various product[4] assortments users find it hard to identify products and services that best fit their wishes and needs. These problems are targeted by various kinds of recommender systems. Collaborative filtering systems manage vast databases of user ratings in order to predict if a user is going to like a certain item or not [15]. Content-based recommenders [20] contain knowledge about the product attributes and try to match it with user preferences. A sub-category of content-based recommenders are knowledge-based recommenders [3]. In addition to knowledge about product attributes and user preferences a knowledge-based recommender (KBR) also manages deep domain knowledge (e.g. relative importance of features and attributes, legislative restrictions, etc.). The following three main tasks for the system in a recommendation cycle can be identified:

1. *Preference elicitation.* The purpose is the collection of user information in order to find out the user's preferences. One possibility to obtain user information is an explicit dialog, where the system poses a number of questions which the user answers in turn. KBRs usually figure out inconsistent user information (e.g. *price = low and type = sports car* in the domain of cars) and offer possible adaptations of the user preferences for the purpose of resolving those inconsistencies [10].
2. *Result calculation.* Based on product features, domain knowledge and information about the user the system calculates utilities of products for the user and thus is able to order the products based on their utility for the specific user. In cases where no suitable products can be found a KBR normally proposes minimal changes of the user preferences which the user can accept in order to obtain a non empty result set.
3. *Result presentation.* After the utility calculation of the products the system typically presents the top ranked products to the user.

Previous research has shown that users typically do not know their own preferences before hand and rather construct their preferences during the recommendation process [13]. This contradicts the strict separation of preference elicitation, result calculation and presentation. One approach to meet this challenge are critique-based recommenders (CBR)[24]. Figure 1 shows a screenshot of a critique-based camera recommender. In a CBR preference construction (instead of preference elicitation) and result presentation happen simultaneously. The system learns user preferences as the user constructs them by interacting with the recommender which, in the case of CBR, means that the user criticizes the presented products (e.g. 'more memory'). After a critique has been placed a new set of products is presented which best matches the critiques the user has stated so far.

Besides preference construction the decision making process of users also needs support. In order to increase a user's confidence in her/his decision it is important that the recommender system applies certain mechanisms to underline differences of the products. Comparison pages where the user can easily compare the features of suitable products are one effective tool to support the user in identifying the most suitable product [11,14]. The CBR shown in Figure 1 also represents a comparison-based approach as the user can always compare

[1] Dept. of Intelligent Systems and Business Informatics,University of Klagenfurt, {alexander.felfernig,stefan.schippel, erich.teppan}@uni-klu.ac.at
[2] Dept. of Interactive Systems, University of Klagenfurt, {gerhard.leitner,rudolf.melcher}@uni-klu.ac.at
[3] Dept. of Cognitive Psychology, University of Klagenfurt, {bartosz..gula,marco.maier}@uni-klu.ac.at
[4] The expressions *product*, *element*, *option*, and *item* will be used synonymously throughout the paper.

three cameras. Although comparison pages already alleviate decision making users often find it hard to take a decision when there are more than one product of similar utility but showing different advantages and disadvantages. In such situations users tend to be dissatisfied with their decision or even quit the purchase process. Previous research has shown that feature-based argumentation and explanations are persuasive techniques which alleviate the decision process [12,19]. The exploitation of context effects [16,27] seems to constitute great potential to push the support of decision making further [17]. Three main application scenarios for such context effects in recommender systems are the following:

- Resolve cognitive dilemmas: In a situation where a decision has to be taken between items of the same utility, the recommender system can give advantage to one of those items and thus lead to a (faster) decision.
- Reinforce confidence: After a decision has been made the recommender system can show the choice set once again adding a decoy element giving advantage to the chosen element.
- Make the user reconsider the decision: In a situation where the user is about to choose a suboptimal product (e.g.: item in the shopping cart), the recommender can ask the user to consider alternatives including the optimal products and corresponding decoys and thus lead to a better decision for the user.

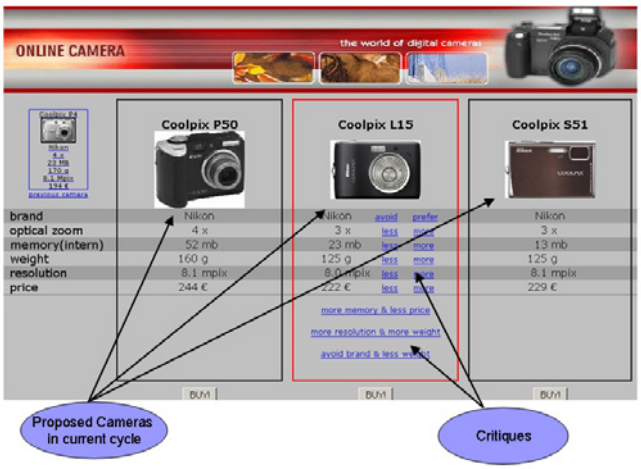

Figure 1. Tweaking-critiquing recommender

There are several types of context effects depending on the relative position of the decoy element (measured in terms of attribute value differences between target, decoy and competitor products, see Figure 2).

A decoy element is an item which is added to a choice set without the purpose of being selected but rather impact on the perception of the other items in the choice set. Typically this leads to an advantage of one (set of) product(s) which we denote as *targets*. Products which do not benefit from the addition of the decoy are called *competitors*. Figure 2 is illustrating the main areas for decoy elements.

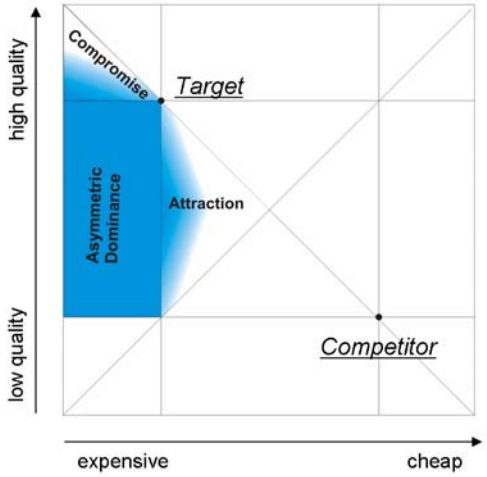

Figure 2. Areas of possible decoys

The *attraction effect* (AE) is occurring when the decoy element's (*D*) attribute values are between the targets' (*T*) and the competitors' (*C*). The major mechanism working behind is called *tradeoff contrast* [28]. Having a target option *T* and decoy option *D* tradeoff contrast is occurring when *T* is much stronger in one (set of) attribute(s) and only a little bit weaker in another (set of) attribute(s) compared to *C*.

without decoy	T	C
Size	21"	19"
Price	400 €	200 €

with decoy	T	D	C
Size	21"	19,5"	19"
Price	400 €	399 €	200 €

Figure 3. Example of Attraction Effect

It is important to note that similarity (distance between products) plays a role in this context. An attraction decoy between *T* and *C* which is more similar to *C* would rather act as decoy for *C* than for *T*. Figure 3 is showing an example with flat screen TVs: After adding a decoy element which is only little cheaper but much smaller the attractiveness of *T* increases.

The *compromise effect* (CE) occurs when the decoy element's (*D*) attribute values are (little) higher on the strong attribute of *T* and much lower on the weak attribute. Here again the major mechanism is *tradeoff contrast*. Figure 4 is showing an example for the *compromise effect*: Again the attraction of *T* is increasing when adding a third TV which is a little bit larger but much more expensive.

without decoy	T	C
Size	21"	19"
Price	400 €	200 €

with decoy	T	D	C
Size	21"	22"	19"
Price	400 €	700 €	200 €

Figure 4. Example of Compromise Effect

For the *asymmetric dominance effect* (ADE) a decoy element is needed which is inferior to *T* in all attributes (i.e. dominated) but compared to *C* it is only inferior in one attribute (i.e. not dominated). Figure 5 is showing an example: *D* is smaller but nevertheless more expensive compared to *T*. Compared to *C* it is larger but more expensive. The classification of dominated and non-dominated alternatives is one reason for the occurrence of the ADE as people tend to apply simpler heuristics than calculating a complete utility function [6]. The other reason is that the ADE is just the special case of the CE and the AE from the point of view of *tradeoff contrast* (see Figure 2).

without decoy	T	C
Size	21"	19"
Price	400 €	200 €

with decoy	T	D	C
Size	21"	20,5"	19"
Price	400 €	410 €	200 €

Figure 5. Example of Asymmetric Dominance Effect

2 THE SIMPLE DOMINANCE MODEL

This section presents a context dependent model for the calculation of the perceived dominance of an item in a given choice set. A dominance value indicates how strong an item appears in the context of other products. Thus, the calculated dominance values give a strong indication on the probability of being selected by a user. The Simple Dominance Model (SDM) focuses on *tradeoff contrast* since it serves as main mechanism for all three main context effects (CE, AE, ADE). As similarity impacts on the power of a decoy element, similarity is also taken into account in the SDM. Furthermore the SDM is multi-preferential since products may be (and usually are) described by more than two attributes and multi-optional because choice sets (like comparison pages) typically consist of more than two products. Another main characteristic of the SDM is its simplicity. This is due to the fact that the less model parameters have to be learned or fine-tuned (in case of definition by an expert) the quicker the model is installed and fit for service.

The core elements of the SDM are dominance values (see Formula 1). A dominance value (DV) expresses how dominant a user perceives a certain item in a given item choice set. Basically, a DV for an item *d* is a weighted ($weight_a$) sum of attribute value differences (a_d-a_i) compared to every other item *i* in the choice set. The value differences are set in relation to the extreme values (max_a-min_a) of the set on a certain attribute. The square root effects that small value differences count relatively more than big differences and thus similarity is taken into account. $Sign(a_d-a_i)$ evaluates to -1 if *d* is worse than *i* (i.e. it d is dominated by *i*) on a certain attribute else it is 1 (i.e. *d* is better or equal).

$$DV_{d \in Items} = \frac{\sum_{i \in \{Items-d\}} \sum_{a \in Attributes} weight_a * \sqrt{\left|\frac{a_d - a_i}{\max_a - \min_a}\right|} * sign(a_d - a_i)}{\#Items - 1}$$

Formula 1. The Simple Dominance Model

Example: Given is a simple choice set consisting of three digital cameras and cameras are described by the two attributes resolution (*MPIX*) and optical zoom (*ZOOM*). The weight for both attributes shall be 0.5[5]. Figure 6 is summarizing the example: The cameras in the set are *T* (9 *MPIX* / 3x *ZOOM*), *C* (3 *MPIX* / 9x *ZOOM*), and *D* (7 *MPIX* / 2x *ZOOM*). *D* represents an asymmetric dominated alternative since *D* is dominated by T but not by *C*.

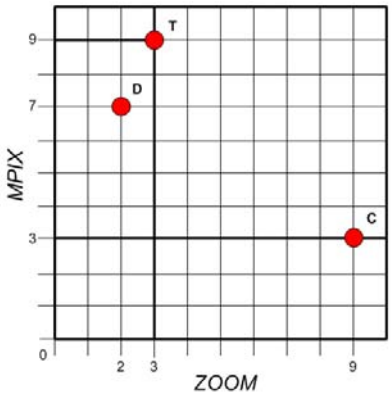

Figure 6. Example set for the calculation of dominance values (DVs)

When we apply Formula 1 the dominance values for *T, D, and C* are calculated as follows:

[5] For simplicity reasons the example choice set only comprises 3 items, 2 attributes and the weights are set to 0.5. However, the model also supports any number of items, attributes, and any weighting. If weights sum up to 1 dominance values are in the range of [-1;1].

$DV_T =$

$$\frac{0.5*\left\langle\sqrt{\frac{9(T;MPIX)-3(C;MPIX)}{9(T;MPIX)-3(C;MPIX)}}+\sqrt{\frac{9(T;MPIX)-7(D;MPIX)}{9(T;MPIX)-3(C;MPIX)}}\right\rangle + 0.5*\left\langle\sqrt{\frac{3(T;ZOOM)-2(D;ZOOM)}{9(C;ZOOM)-2(D;ZOOM)}}-\sqrt{\frac{9(C;ZOOM)-3(T;ZOOM)}{9(C;ZOOM)-2(D;ZOOM)}}\right\rangle}{2}$$

$=0.25$

$DV_D =$

$$\frac{0.5*\left\langle\sqrt{\frac{7(D;MPIX)-3(C;MPIX)}{9(T;MPIX)-3(C;MPIX)}}-\sqrt{\frac{9(T;MPIX)-7(D;MPIX)}{9(T;MPIX)-3(C;MPIX)}}\right\rangle + 0.5*\left\langle-\sqrt{\frac{3(T;ZOOM)-2(D;ZOOM)}{9(C;ZOOM)-2(D;ZOOM)}}-\sqrt{\frac{9(C;ZOOM)-2(D;ZOOM)}{9(C;ZOOM)-2(D;ZOOM)}}\right\rangle}{2}$$

$=-0.275$

$DV_C =$

$$\frac{0.5*\left\langle-\sqrt{\frac{9(T;MPIX)-3(C;MPIX)}{9(T;MPIX)-3(C;MPIX)}}-\sqrt{\frac{7(D;MPIX)-3(C;MPIX)}{9(T;MPIX)-3(C;MPIX)}}\right\rangle + 0.5*\left\langle\sqrt{\frac{9(C;ZOOM)-3(T;ZOOM)}{9(C;ZOOM)-2(D;ZOOM)}}+\sqrt{\frac{9(C;ZOOM)-2(D;ZOOM)}{9(C;ZOOM)-2(D;ZOOM)}}\right\rangle}{2}$$

$=0.025$

In a set only consisting of *T* and *C* both would have a DV of zero. Thus, the addition of *D* shows a clear impact on *T*'s dominance. A DV of -1 for an item would mean to be fully dominated, i.e. to be the worst product on every attribute. A DV of 1 for an item would mean that the product is best on every attribute.

One major application of the SDM is to detect decoy items for a certain target item in a given result set in order to break up the cognitive dilemma. Decoy items in this case are simply items which maximize the dominance value of a target item when added to the result set. Usually, the knowledge base of a recommender system comprises hundreds or even thousands of items. After result calculation typically a small set remains to make up the result set. Commonly the top ranked (i.e. highest utility) items are presented to the user. A cognitive dilemma can now occur when there are multiple items of similar (or equal) utility. One approach to break up the dilemma is to give advantage to one of those "similar" items. This may be the item with the highest utility (in case of similar utilities) or the best item on a specific attribute (e.g. in case of equal utility give advantage to the item with the lowest price). Depending on the result set's size there may be several dilemmas. The task is to find decoys for all target items, i.e. items which should be preferred. The procedure of finding the optimal decoys for specific target items is shown in Algorithm 1.

```
procedure decoy (Items[] topItems, Items[] targets,
  Items[] possibleDecoys) -> Item{
    Number current=
          sumOfDominanceValues(targets, topItems);
    Item[] decoys;
     for each(x IN  possibleDecoys){
       if(sumOfDominanceValues(targets, topItems+x)
          > current){
        current= sumOfDominanceValues (target, topItems+x);
        decoy.add(x);
       }
   return decoy;
}
```

Algorithm 1. Procedure for finding the ideal decoy element(s)

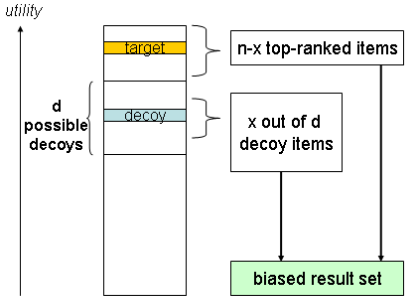

Figure 7. Principle of decoy algorithm

The basic principle of the procedure is visualized in Figure 7. If the result set is consisting of n items then n-x items are those which are top ranked by the recommender system. The x decoy items are determined out of d (range is defined by domain expert) possible decoys which are near the top ranked items. This has the following reason: since it is also possible that consumers choose the decoy element the utility of the decoy must not be too low in order to guarantee a good recommendation. In other words even if the consumer chooses the decoy this should be a good recommendation. For example, if a result set's size is 5 and there are 10 products which are not filtered out by the recommender system then the result set might consist of the three top-ranked items + two decoys out of the rest. It is important to note that the application of the SDM happens after the recommender has filtered out completely unsuitable items (e.g.: items exceeding a minimum/maximum threshold on a certain attribute).

The SDM's parameters are simply the attribute weightings of Formula 1. In order to set up the SDM its parameters have to be defined by domain experts or learned from statistical data (or a combination). At this point the question about user specific versus uniform weights for all users arises. Of course, user specific weights guarantee optimal performance of the SDM for all users but there are reasons why recommendation processes often lack user specific weights:

- Many recommendation domains are not subject of frequent purchase behaviour like cars, digital cameras, financial products, etc, which leads to an absence of long-term statistical information.
- Anonymous use of recommenders avoids the use of long-term user profiles.
- Even if long-term user profiles are available, such profiles tend to be sparse [21].

- The amount of information which can be elicited during a recommendation process is limited as the amount of time users spend with the system is limited.

So in many cases uniform weights would at least serve as a basis for the application of the SDM. If there is good learning data available the procedure in Algorithm 2 can be applied to train the model: the basic idea is simply to find the attribute weightings with which the model predicts the most correct choice outcomes, i.e. where the selected product gets the highest dominance value. Which algorithm has to be applied to obtain the set of *possibleWeightings* (see Algorithm 2) depends on the amount of learning data. If there is not much learning data available a brute force method where all weights from zero to one on every dimension are tested can be sufficient. To manage a large amount of learning data more intelligent algorithms like some sort of heuristic search has to be applied.

```
procedure learnWeights(ChoiceSet[] choiceSets){
  Weighting weighting;
  Number hit=0;
  for each(w IN possibleWeightings){
    number current=0;
    for each(set IN choiceSets){
          if(max(dominanceValues(set))==
                    domianceValue(selected,set)){
                              current=current+1;
                    }

    }
    If(current>hit){
        Hit=current;
        weighting=w;
    }
  }
return weighting;
}
```

Algorithm 2. Procedure for learning weights from statistical data

3 MODEL PERFORMANCE

A supervised user study with the goal of testing the predictive power of the presented model was carried out with students and staff of the Klagenfurt University. Altogether we had 37 participants (18 females) with a mean age of 26.08 (SD=7.21, range: 19 - 60) which interacted with a critique-based digital camera recommender (Figure 1). In the end of the interaction the system proposed three cameras which the participants had to choose from ('buy'). The proposed cameras where taken out of a product base which consisted of 231 different cameras and corresponded to individual preferences specified during the recommendation process. Table 1 summarizes the 37 choice situations. Details of the presented cameras in the choice sets are shown in Table 2. The SDM was used to calculate dominance values and thus allow predictions on the participants' choices. The predictive power of the model is the core factor in order to calculate suitable decoy elements. The procedure of model application followed the general logic of a leave-one-out cross-validation[6] [9] and was as follows:

1. Take 36 out of 37 datasets and learn weights (i.e. apply Algorithm 2 in a brute force manor)[7].
2. With the learned weights calculate dominance values for the 37th dataset.
3. Mark as hit, if selected camera has the highest DV.
4. Repeat for all 37 datasets.

choiceSet	IdCam1	Dom1	IdCam2	Dom2	IdCam3	Dom3	selectedCam
1	78	-0,482	81	0,968	82	-0,486	78
2	151	0,422	143	-0,335	158	-0,086	143
3	38	-0,261	49	0,498	34	-0,236	38
4	215	-0,424	97	0,306	49	0,118	97
5	159	-0,187	153	0,317	222	-0,130	222
6	206	-0,671	38	0,098	223	0,573	223
7	202	0,102	200	0,230	129	-0,332	200
8	99	-0,799	96	0,483	103	0,316	96
9	223	0,582	206	-0,609	104	0,026	223
10	80	-0,116	81	-0,055	200	0,172	81
11	68	-0,328	67	0,304	135	0,024	67
12	175	0,887	174	-0,551	173	-0,336	175
13	50	0,208	46	0,334	43	-0,541	46
14	197	0,138	196	-0,141	203	0,003	197
15	125	0,016	72	0,251	99	-0,266	99
16	149	-0,154	152	-0,011	150	0,165	150
17	154	0,323	56	-0,033	137	-0,290	56
18	148	0,230	145	-0,765	144	0,536	144
19	201	0,217	200	0,619	194	-0,836	201
20	176	-0,133	168	-0,107	167	0,240	176
21	215	-0,182	213	-0,353	211	0,536	215
22	203	-0,064	202	0,564	215	-0,500	202
23	132	0,059	134	0,052	131	-0,112	132
24	42	-0,050	44	0,118	45	-0,068	42
25	83	0,501	85	-0,093	84	-0,408	83
26	121	-0,346	126	0,286	120	0,060	126
27	127	1,067	153	0,140	206	-1,208	127
28	188	0,367	176	-0,029	168	-0,338	188
29	158	-0,584	151	0,086	150	0,498	158
30	80	0,527	79	-0,767	76	0,241	80
31	205	0,137	206	0,322	217	-0,459	205
32	7	-0,031	8	0,016	9	0,016	9
33	17	-0,878	127	0,975	122	-0,097	127
34	201	-0,042	199	0,229	215	-0,187	215
35	85	-0,093	83	0,501	84	-0,408	84
36	214	-0,040	231	0,080	229	-0,040	231
37	154	-0,407	211	0,596	51	-0,188	211

Table 1. Dominance-based prediction (Dom1-3 = calculated dominance values for the presented cameras)

Altogether 22 out of 37 choices were predicted correctly (highlighted in Table 1), which makes up a ~60% hit rate with a 95%-confidence-interval ranging from 42.9% to 76.1%. The choice of target cameras predicted by the model differed significantly from 1/3 chance expectation (One-Sample t-Test; $t(36)=3.193$, $p=.001$, one-sided). The performance of the model is highly related to the correctness of the weights, even though we have found that also a small amount of learning data (in our

[6] The reason of applying a leave-one-out method was the small amount of learning data. If enough learning data is available the typical approach is to partition the set of data into a set of learning data and a set of test data.

[7] The weights only varied very slightly. The ranking was as follows:

1. price [> 30%]
2. optical zoom [> 25%]
3. resolution [> 20]
4. memory [> 10%]
5. weight [< 10%]

case 36 out of 37 datasets) produces reasonable results. The predictive value was quite remarkable, as the experiment was carried out under real world conditions using a real recommendation environment. Three important factors were weakening the predictive power, as they cannot yet be taken into account in the model:

- Cameras were additionally represented by images, which can (and typically do) influence choice [31].
- The attribute 'manufacturer', although shown in the experiment, cannot be taken into account by the model, as it is no ordinal attribute (but for creating a realistic recommendation situation it was necessary to show this attribute).
- The last set (the participants had to choose from) consisted of three similar cameras as after every interaction with the recommender (i.e. critiquing one of the current cameras) three adequate and similar cameras were presented to the participant. It is evident that the more similar the items are the more difficult it is to predict the choice outcome.

4 RELATED WORK

A lot of work in the area of context effects has been conducted since the seventies [1,2,4,5,7,8,16,18,23,25,26,30]. Although most of the research has been done to show context effects in different application domains [4,23,26,30] some research has also be done trying to find a model for context effects.

Models for single effects. Most of this research was concentrating on single effects. Models which focused on the similarity effect are found in [1,7]. Models trying to explain the attraction effect are found in [5,8]. A good overview of diverse models targeting the compromise effect is found in [16].

Models incorporating multiple context effects. One approach of taking into account more than one single effect was the application of neural networks. [18] presents a neural network for choice behaviour but lacks the multi-optionality (choice set > 2). A feed forward neural network considering the asymmetric dominance effect, attraction effect, and compromise effect is found in [25]. The problem with neural networks in general is the cold start problem. Whereas most mathematical models provide reasonable (though not optimal) results using standard parameters (often defined by domain experts), the application of neural networks without time consuming training based on good training data is most of the time not possible. Thus, what is needed is a mathematical model which is multi-preferential (number of attributes > 2), multi-optional (size of choice set > 2), and accounts for multiple context effects. A complete mathematical model is presented in [2]. The main difference to the model presented in this paper is the complexity and the number of parameters. The set of parameters of the Simple Dominance Model (SDM) only consists of the product attribute weightings. The main reason why the SDM manages to stay very simple is that instead of modelling various effects the SDM rather concentrates on one important common factor of the main context effects which is *tradeoff contrast*.

5 ETHICAL ISSUES

A certain challenge which is coming along with persuasive technologies in general is that to some extend persuasive techniques could be exploited not only to have positive impacts but also to negatively influence a certain involved party. In the context of online selling systems like recommenders the biggest danger is that sellers might apply persuasive techniques in order to increase profit without taking too much care about the users' wishes and needs and thus would rather manipulate than recommend. The major reason why the application of the SDM can be seen as less problematic is that the application of the SDM happens well after the recommender system already has sorted out unsuitable products. In cases where no suitable products can be found, knowledge-based recommenders typically inform the user about the empty result set and offer repair mechanisms in order to support the user in adapting the preferences. Either way the SDM is only applied on the basis of a set of suitable items. Furthermore, the application of the SDM should not prevent the top-ranked items to be displayed but rather underline differences of those top-ranked items (see Algorithm 1 in Section 2). Moreover it is very substantial to emphasize that as soon as more than one product is presented at the same time context effects are occurring which leads to the conclusion that there are basically two possibilities for dealing with the existence of context effects. The first possibility is to simply ignore the existence of those effects. The second, more constructive approach is to analyze those effects and to develop tools (like the SDM) in order to control those effects and identify potentials for effective user support.

6 CONCLUSIONS & FUTURE WORK

This paper presents the Simple Dominance Model which can be used to calculate context dependent dominance values for items in a choice set. Context dependent dominance values can be used for finding optimal decoy items in order to decrease the cognitive choice dilemma in a choice situation of recommender systems, increase the user's confidence in its own decision, and persuade the user to reconsider its decision and thus lead to a better choice. The model is multi-preferential (i.e. more than two product attributes are supported) and multi-optional (i.e. more than two products in a choice set are considered). A first proof of concept has been provided by showing the predictive power of the model by application on empirical data gained from a user study in the domain of recommender systems. The simplicity of the model is one of its main characteristics. The only parameters to be set are product attribute weightings. An algorithm for learning weights in this context is presented. The future work consists mainly of testing the model's performance in different recommendation domains with various numbers of attributes and various sizes of choice sets. Furthermore, the performance will be compared to more complex models targeting similar purposes.

id	zoom	res	price	mem	weight	id	zoom	res	price	mem	weight	id	zoom	res	price	mem	weight	id	zoom	res	price	mem	weight
7	0	3	61	16	130	80	10	7,1	162	27	306	134	12	8,2	242	24	340	188	5	7,1	179	20	115
8	0	5	91	16	140	81	10	8	207	26	306	135	5	8,1	147	32	161	194	10	6	318	10	285
9	0	5	91	16	140	81	10	8	207	26	306	137	5	6,2	199	10	170	196	18	7,1	320	20	365
17	3	10,1	180	24	140	82	10	6,3	316	0	660	143	3	8	222	23	125	197	18	8	384	47	365
34	4	7,1	312	0	150	83	18	8	339	58	410	144	3	6	119	23	120	199	12	6	319	32	389
38	4	12,1	391	0	165	84	10	9	449	0	645	145	3	5,1	179	23	120	200	10	8,1	238	31	380
42	4	5	107	0	165	85	10	9	399	0	650	148	3	6	145	23	125	201	12	7,2	349	32	406
43	4	6	199	0	180	96	3	8	156	32	180	149	4	8	298	32	170	202	15	8,1	296	31	540
44	4	7,1	134	0	160	97	6	8	176	16	170	150	4	8,1	194	23	170	203	15	8,1	331	31	540
45	4	7,1	140	0	165	99	3	6,2	283	32	140	151	4	8,1	244	52	160	205	3	8,1	354	26	155
46	4	7,1	175	0	175	103	3	8	175	32	200	152	4	10	314	21	200	206	3	10,1	379	26	155
49	6	12,1	371	0	300	104	3	8,2	255	32	170	153	4	12,1	345	52	200	211	6	8	206	54	185
50	6	7,1	258	0	210	120	3	8,2	143	16	115	154	10	6	221	16	220	213	5	8,1	352	31	172
51	6	8	216	0	200	121	3	8	144	32	110	158	3	8,1	229	13	125	214	3	8,1	228	31	159
56	7	8,1	282	11	149	122	3	10,1	150	32	142	159	3	8,1	263	52	125	215	5	8,1	329	31	186
67	4	8,3	121	10	155	125	3	6,1	233	32	120	167	5	7,1	259	17	114	217	3	7,2	430	58	132
68	4	8,3	149	10	155	126	3	8	130	32	142	168	5	7,1	276	17	120	222	3	8,1	294	64	161
72	3	6,3	239	26	155	127	5	12,1	167	64	161	173	3	8	239	19	181	223	3	12,1	258	31	142
76	4	8,2	156	12	140	129	10	6,1	258	32	287	174	3	8	279	28	145	229	3	7,2	179	31	125
78	5	6	259	32	195	131	10	7,1	169	32	285	175	5	8	209	47	125	231	3	8,1	224	31	155
79	10	5,1	198	0	410	132	12	7,1	211	32	300	176	5	8	291	15	125						

Table 2. Details of presented cameras in the choice sets

REFERENCES

[1] D. Ariely, T. Wallsten, Seeking subjective dominance in multidimensional space: Anexploration of the asymmetricdominance effect, Organizational Behaviour and Human Decision Processes, 63(3), 223-232, 1995.

[2] M. Bhargava, J. Kim, R. Srivastava, Explaining Context Effects on Choice Using a Model of Comparative Judgment, Journal of Consumer Psychology, 9(3), 167-177, 2000.

[3] R. Burke, Knowledge-based Recommender Systems. Encyclopedia of Lib. and Information Syst., 69(32), 180-200, 2000.

[4] S. Callander,C. H.Wilson, Context-dependent Voting, Quarterly Journal of Political Science, 227–254, 2006

[5] M. Candel, A probabilistic feature model for unfolding tested for perfect and imperfect nestings, Journal of Mathematical Psychology, 41(4), 414-430, 1997.

[6] M. Devetag, From utilities to mental models: A critical survey on decision rules and cognition in consumer choice, CEEL Working Papers 9902, Computable and Experimental Economics Laboratory, Department of Economics, University of Trento, Italia, 1999.

[7] R. Dhar, R. Glazer, Similarity in context: Cognitive representations and violations of preference and perceptual invariance in consumer choice, Organizational Behaviour and Human Decision Processes, 67, 280-293, 1996.

[8] S. Edgell, W. Geisler, Aset theoretic random utility model of choice behaviour, Journal of Mathematical Psychology, 21(3),265-278, 1980.

[9] B. Efron, R. J. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.

[10] A. Felfernig, G. Friedrich, D. Jannach, M. Zanker, An Environment for the Development of Knowledge-based Recommender Applications, International Journal of Electronic Commerce (IJEC), 11(2), 11-34, 2006.

[11] A. Felfernig, B. Gula, An Empirical Study on Consumer Behaviour in the Interaction with Knowledge-based Recommender Applications, IEEE Joint Conference on E-Commerce Technology (CEC06) and Enterprise Computing, E-Commerce and E-Services (EEE06), San Francisco, California: IEEE Computer Society, 288–296 , 2006.

[12] A. Felfernig, B. Gula, G. Leitner, M. Maier, R. Melcher, E.C. Teppan, Persuasion in Knowledge-based Recommendation, to appear in proceedings of Persuasive Technologies 2008, Oulu, Finnland, 2008.

[13] G. Häubl, K.B. Murray, Processes of preference construction in agent-assisted online shopping, Online consumer psychology: Understanding and influencing behavior in the virtual world, ISBN-10: 0805851550, Lawrence Erlbaum Associates Inc, 265-286, 2005.

[14] G. Häubl, V. Trifts, Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids, Online consumer psychology: Understanding and influencing behavior in the virtual world Marketing Science (19), Special Issue on Marketing Science and the Internet, 4-21, 2000.

[15] J. L. Herlocker, J. A. Konstan, J.Riedl, Explaining Collaborative Filtering Recommendations. Procs. ACM Conf. on CSCW. Pennsylvania, USA, 2000.

[16] R. Kivetz, O. Netzer, V. Srinivasan, Alternative Models for Capturing the Compromise Effect, Stanford University, Stanford GSB Research Paper No. 1731(R) , 2003.

[17] N. Klein, M. Yadav, Context Effects on Effort and Accuracy in Choice: An Enquiry into Adaptive Decision Making The Journal of Consumer Research, (15/ 4), 411-421, 1989.

[18] S. Leven, D. Levine, Multiattribute Decision Making in Context: A Dynamic Neural Network Methodology, Cognitive Science, 1996.

[19] J. Masthoff, N. Tintarev, Effective explanations of recommendations: user-centered design, ACM conference on recommender systems, 153-156, 2007.

[20] M. Pazzani,D. Billsus. Learning and Revising User Profiles: The Identification of Interesting Web Sites. Machine Learning, (27):313–331, 1997.

[21] M. Papagelis,D. Plexousakis, T. Kutsuras, Alleviating the Sparsity Problem of Collaborative Filtering using Trust Interfaces, Trust Management, ISBN: 978-3-540-26042-4, Springer Lecture Notes in Computer Science, 2005.

[22] H. Pechtl, Definitions- und Wirkungsbereiche des decoy-Effekts eine empirisch-explorative Untersuchung, Nov. 2004, Diskussionspapier 10/04 ISSN 1437 – 6989.

[23] M. Ouyang, Does the Decoy Effect Exist in the Marketplace? -- An Examination of the Compromise Effect, Congrés 2004 de l'Association des Sciences Administrative du Canada, 2004.

[24] J. Reilly, K. McCarthy, L. McGinty, . Smyth, Incremental Critiquing, Knowledge Based Systems, 18(2-3), 143 – 151, 2005.

[25] R. R. Roe, J. Busemeyer, T. Townsend, Multi-alternative decision field theory: A dynamic artificial neural network model of decision making, Indiana university, 1999.

[26] M. Schweizer, Kontrast- und Kompromisseffekt im Recht am Beispiel der lebenslänglichen Verwahrung, Schweizerische Zeitschrift für Strafrecht, 123 (4)438-457, 2005.

[27] I. Simonson, Choice Based on Reasons: The Case of Attraction and Compromise Effects, Journal of Consumer Research (16), 1989.

[28] I. Simonson,A. Tversky, Choice in context: Tradeoff contrast and extremeness aversion, in: Journal of Marketing Research (39), 281-292, 1992.

[29] I. Simonson, A. Tversky, Context-Dependent Preferences, Management Science, 39(10), 1179-1189, 1993.

[30] C. Tan, Y. Chan, X. Yang, H. Chan, H. Teo, Effects of Choice Contrast and Order Sequence on Consumer Judgement and Decision

in Comparison-Shopping Assisted Environment, Third Annual Workshop on HCI Research in MIS, 2004.
[31] R. Westfall, Psychological Factors in Predicting Product Choice, Journal of Marketing, 26(2), 1962.

Persuasion Technology Through Mechanical Sophistry

Micah Clark and **Selmer Bringsjord**[1]

Abstract. We introduce an approach to digital persuasive technology. In this approach, computing machines are given the power to persuade by generating what we call *illusory arguments*. These arguments derive their power in significant part by exploiting the empirical fact that humans succumb to *cognitive illusions*. The persuasion technology we present can be used for weal or woe, but at least for now, we approach the subject purely as AI engineers, not ethicists.

1 INTRODUCTION

AI has developed systems involving argumentation in two ways: modeling *with* argument, and modeling *of* argument [26]. Our work is in neither of these categories, but it is inspired by the distinction. We aim to engineer a machine that generates sophistic arguments in order to persuade. Such a machine could be used to persuade humans of truth or falsehood, and our emphasis is on the latter: persuading humans of falsehoods. As an engineering endeavor, constructing a machine that *lies* via sophistic argument makes sense, because, if the engineering succeeds for falsehoods, it can be adapted to persuade humans of truth.

We introduce a psychology of reasoning-based approach to mechanical sophistry, namely an approach for generating *illusory arguments* — arguments that compel erroneous belief by exploiting the dichotomy of what humans *ought* to believe and what humans *will* likely believe. In illusory arguments, the gap between *ought* and *will* is significant and predictable; these arguments incorporate, and prey upon, reliable systematic errors in human reasoning. In short, illusory arguments are psychologically persuasive yet fallacious arguments that lead humans into erroneous beliefs while simultaneously confirming their naïve intuitions about what is true.

In the mechanical sophistry we propose, illusory arguments emerge from an ecumenical theory of human reasoning. This paper roughly circumscribes our approach to sophistry; an approach novel in its application of empirically supported psychological theories to the practical challenge of persuading humans through argument.

2 SOPHISTIC PERSUASION

Some have viewed persuasion in argument as what force or influence a justification should have (surveyed, e.g., in [5, 23]), using as their lens a continuum of strength factors (e.g., Chisholm's [6]) and frameworks for defeasible reasoning in the tradition of Pollock [21] and Toulmin [31]. Others have treated persuasion in argument as a way to resolve a difference of opinion, or to establish a common belief or intent [32, 17], studying persuasion's game-theoretic properties in multi-agent dialog systems (surveyed in [22]). There seem to be few studying persuasion (e.g., [10, 27]) who implement machines capable of persuading humans.

[1] Rensselaer Polytechnic Institute, USA, email: {clarkm5, brings}@rpi.edu

Our work appeals to an older, decidedly philosophical conception of persuasion: sophistry. The essence of sophistry *is* persuasion, the intentional manipulation of others' beliefs through the force of argument.[2] The skilled practitioners of sophistry are those whose rhetorical methods successfully manipulate.

There are two broad sorts of persuasive arguments, "the one producing belief without knowledge, the other knowledge" [19]. Likewise, there are two archetypal practitioners: the *liar* and the *truth-teller*. The truth-teller argues in good faith for what he/she believes with the intention of producing knowledge. The liar argues falsely, and sophistically, for what he/she does not believe with the intention of producing belief without knowledge.[3] Of the two, the successful liar demonstrates greater skill than the successful truth-teller: the truth-teller has in persuasion an ally, truth, while the liar must overcome, and conquer, truth as an adversary. It is this inequity of skill that calls us to study arguments that deceive rather than arguments that inform, and liars rather than truth-tellers.

3 COGNITIVE ILLUSIONS

Our theory of sophistry is driven by the difference between what humans *will* likely believe in light of arguments, and what ideally rational agents *ought* to believe in light of said arguments. According to the neo-Piagetian view of the development of bias-free human reasoning: Given sufficient training, neurobiologically normal humans can reason in a normatively correct, bias-free manner [3, 28]. A mounting body of empirical evidence of *cognitive illusions* (see, e.g., [18, 20, 1]) strongly suggests the inverse of the neo-Piagetian view is also true, that being: Naïve humans (i.e., those untrained or insufficiently trained in logic) often reason in an incorrect, bias-plagued manner, and they do so in regular, systematic fashion.

Cognitive illusions designate phenomena where the beliefs and behavior of humans are consistent and predictable, yet deviate substantially from what is ideally rational. Worse still, when succumbing to cognitive illusions humans believe, and report, that they are thinking and acting rationally. The allusion here to "illusions" is not accidental. It reflects the position that formal, normative rationality has primacy over base human intuition.[4] Even if one rejects formal rationality as the appropriate benchmark for human rationality, the fact remains that deviation from the maxims of formal rationality leads inexorably to the "Dutch book" [25, 8], i.e., to situations where one will believe falsehoods and act against one's material self-interest. Cognitive illusions are, in essence, empirically validated exemplars of human vulnerability to "Dutch book" situations.

[2] "[Sophistry] is the creator of persuasion, and that all its activity is concerned with this, and this is its sum and substance." [19]

[3] There are many variants of the two archetypes; e.g., one can speak a falsehood believing it to be true, and a fallacious argument can be given in support of a true conclusion.

[4] This position is dominate but not universal (see, e.g., [4], for discussion).

4 ILLUSORY INFERENCES

Illusory inference is a particular class of cognitive illusion wherein naïve humans, when presented with a collection Φ of declarative sentences, overwhelmingly conclude ψ, though ψ does not follow from Φ in any standard proof calculus c (i.e., $\Phi \nvdash_c \psi$). One example is:

> "All the Frenchmen in the room are wine-drinkers. Some of the wine-drinkers in the room are gourmets." [11]

When asked what, if anything, they can rightly conclude from the above sentences, the majority of human subjects erroneously infer that some of the Frenchmen in the room are gourmets, while in fact, no relation between Frenchmen and gourmets can be validly inferred. Another example (viz., the *king–ace* illusion) is:

> "If there is a king in the hand, then there is an ace, or if there isn't a king in the hand, then there is an ace, but not both of these if-thens are true." [11]

When asked about this sentence, almost every subject draws the conclusion that there is an ace in the hand, even though it follows deductively that there *cannot* be an ace in the hand. To see why there cannot be an ace, recognize that one of the two conditionals in the premise must be false, and a conditional is false only when its antecedent is true and its consequent is false. If "if there is a king in the hand, then there is an ace" is false, then there is not an ace, and if "if there isn't a king in the hand, then there is an ace" is false, then again, there is not an ace. So regardless of which conditional is false, there cannot be an ace in the hand. If the subjects had been offered wagers on the correctness of their answers, they would have been summarily fleeced.

How humans make inferences and draw conclusions (correct or otherwise) is, in psychology, a matter of controversy. Some hold that humans reason by application of inference schemata akin to a proof theory, i.e., a sort of "mental logic" [29, 2]. But explaining illusory inferences has challenged the proponents of mental logics.[5] In contrast, illusory inferences are consonant with the view that humans reason by constructing and inspecting iconic representations homomorphic to real or abstract situations, so called "mental models" [12, 30]. According to mental model theories, illusory inferences are the result of incomplete mental models; in response to limited cognitive resources, humans normally construct only partial mental representations. These partial representations are usually sufficient for correct reasoning but, under certain conditions, they lead to compelling yet erroneous conclusions, i.e., illusory inferences. Mental model theories have successfully predicted illusions in, e.g., modal, deontic, spacial, causal, and probabilistic reasoning domains.

Human reasoning is heterogeneous. Our ecumenical theory subsumes elements and ideas from mental logics and mental model theories. The bridging of mental logics and mental models is an aspiration our theory shares with *mental meta-logic* [33, 34]; however, their pursuit and ours are quite different. Mental meta-logic pursues a predictive, statistical theory of perceived difficulty and objective accuracy in reasoning tasks, while we pursue predictive simulation of human reasoning through a unified inferential calculus. Toward this end, we developed an inferential calculus modeling human sentential reasoning. By then varying the definitions of the calculus' basic operators we formed a lattice of calculi, each modeling a degree of reasoning competency distinguished according to completeness of mental representation, and ability to compensate for remaining incompleteness. The calculi share a decision procedure for assessing consistency, validity, and modal consequence (i.e., possible consequence, necessary consequence); the actual assessments (e.g., what sentences are validities), of course, differ between calculi.

Space constraints do not permit us to detail the calculi and decision procedure here, but their import is this: Through a short series of diagnostic questions we can identify the specific calculus that best approximates an individual subject's reasoning ability, and thus make subject-specific predictions of *perceived* consistency, validity and (modal) consequence in sentential reasoning.

5 ILLUSORY ARGUMENTS

Plato held that sophistry could only sway the uneducated and uninformed.[6] But cognitive illusions, and illusory inferences in particular, suggest that Plato was wrong. There might be a sophistic technique persuasive to domain experts and the ignorant alike; a technique based on illusory inferences to which few are immune, viz., *illusory arguments*.

To explain, theories of human reasoning (including our own) are generally extrapolations from the results of empirical studies wherein subjects are asked to perform some constructive reasoning task.[7] These tasks are "constructive" in the sense that subjects are tacitly expected to originate a line of reasoning, or justification, for their answers. Illusory inferences are also studied in this fashion. What has been largely ignored are evaluative tasks, where subjects are asked to agree or disagree with a determinate line of reasoning, i.e., to accept or reject an argument given to them.

The hypothesis of *illusory arguments* is this: If humans regularly err in concluding that ψ follows from a set Φ of declarative sentences due to their own imperfect reasoning, they should all the more affirm ψ when presented with an argument mirroring what would likely be their own flawed justification for ψ. For example, many may be persuaded of the presence of an ace in the hand (in the *king–ace* illusion) by the following illusory argument:

> *We know that either there is a king in the hand or there isn't. Suppose that there is a king. Now, we also know that if there is, then there is an ace. We know this from the conditional that "if there is a king in the hand, then there is an ace." Therefore, there is an ace in the hand.*
>
> *Now, let's suppose that there* isn't *a king in the hand. This connects to the other if-then: "if there isn't a king in the hand, then there is an ace." Therefore, we can infer again that there is an ace in the hand.*
>
> *So, while we don't know whether there's a king in the hand or not, it doesn't matter: There is an ace in the hand.*

We represent an argument $\mathcal{A}$ as a finite sequence of inferences much as one might articulate a natural deduction proof. The semantic

[5] The inference schemes and procedures associated with mental logics are typically valid yet incomplete — all drawn conclusions are valid but not all valid conclusions are reachable. Yet in illusory inferences, the conclusion drawn is invalid. Proponents of mental logics explain illusory inferences as inadvertent deviations from the prescribed process, e.g., mis-applying, or failing to apply, an inference schema. This style of explanation has largely precluded mental logics from predicting illusory inferences.

[6] "[Sophistic rhetoricians] merely discover a technique of persuasion, so as to appear among the ignorant to have more knowledge than the expert [...] for surely, among those who know, he will not be more convincing than the [expert]." [19]

[7] E.g., to decide what can be concluded from premises, to decide whether one premise is possible given the truth of another, to describe what would make a premise false, etc.

structure of an argument is "natural" insofar as it includes the basic format and structures of Fitch-style [9] natural deduction extended to informal arguments. In a proof, each sentence s_i must follow from one or more of the preceding $i-1$ sentences in accordance with the specific proof theory at hand. The proof theory also determines the "size" of an inferential leap. In illusory arguments, the ith sentence need only appear to follow from what precedes it, and the size of an inferential leap may be arbitrary so long as the veneer of logical consequence is maintained. With regard to the inferential calculi mentioned in §4, an inference is *perceived* as valid if its conclusion is perceived to be a necessary consequence of its premises, and an argument $\mathcal{A}$ is *perceived* as valid if all of its inferences are perceived as valid. Thus, the calculi's decision procedure for (modal) consequence can be extended to decide the validity of inferences and arguments.

In illusory arguments and sophistry, veracity is not our concern, instead credibility (i.e., believability, persuasive force) is. We use as the measure of an argument's credibility, its probability of acceptance by a naïve human or humans. Based on prior published results (and our own ongoing experiments) we are developing a parametric model of credibility for our theory of human sentential reasoning. In its current embryonic form, our model equates the frequency at which particular inferences are made in constructive tasks to the probability that those inferences will be accepted in evaluative tasks. For example, the probability of acceptance (i.e., credibility) for *affirming the consequent* is assumed to be the observed frequency at which subjects commit this fallacy in constructive tasks. The credibility of an argument $\mathcal{A}$ is taken to be the probability that all necessary inferences, leading from initial premises to final conclusion, are accepted.

This model of credibility is quite impoverished; we do not expect it to predict actual argument acceptance rates. Fortunately, in sophistry objective accuracy is not required. All that is needed is a relative ordering of arguments, i.e., a determination of whether one argument is more, or less, credible than another — though obviously, the more accurate one's prediction of credibility, the better one's discrimination between arguments. We are working toward a more sophisticated model of credibility, one that makes reasonably accurate predictions of argument acceptance rates. However, there are significant difficulties in extracting such a model from extant studies of performance in constructive tasks; we briefly mention two of them.

1. We are interested in making single-subject predictions of argument credibility, yet most of the relevant studies report only aggregate subject data.
2. Core competence may subvene the reasoning in both constructive and evaluative tasks, but it seems unlikely that the strategies used to direct reason would be common across these tasks. For example, subjects on their own rarely infer a disjunctive statement from a categorical assertion (i.e., disjunction-introduction), but when evaluating the reasoning of another, it is doubtful that they would so rarely recognize the validity of this inference.

These two difficulties (as well as others) point out that further study of human reasoning in evaluative tasks is needed if we are to achieve the goal of a truly predictive model of credibility.

6 MECHANICAL SOPHISTRY

Our approach to sophistry is made concrete in the machine $\mathcal{M}$. $\mathcal{M}$ is a normatively correct reasoner, i.e., $\mathcal{M}$'s reasoning is valid; it employs a standard inferencing scheme for many-sorted logic [15] (we use $\vdash_{MSL}$). $\mathcal{M}$ reasons over its own beliefs, and its beliefs about the beliefs of $\mathcal{D}$, whom $\mathcal{M}$ intends to deceive or persuade. Using capital Greek letters Γ and Σ to indicate sets of declarative sentences, we say $\mathcal{M}$ believes Γ, and $\mathcal{M}$ believes that $\mathcal{D}$ believes Σ. Further, $\mathcal{M}$ believes that $\mathcal{D}$'s reasoning capacity and process accords with our ecumenical theory of human reasoning, i.e., $\mathcal{M}$ has a *theory of mind* (ToM) [24] relative to $\mathcal{D}$ (we use $\vdash_{ToM}$). $\mathcal{M}$ does not assume that $\mathcal{D}$ is logically omniscient with respect to the ToM. Rather, the ToM describes what $\mathcal{D}$ is likely to believe upon sufficient reflection; that is to say, the ToM describes $\mathcal{D}$'s *implicit beliefs* (see, e.g., [14, 13]).

In this framework, $\mathcal{M}$ can deceive $\mathcal{D}$ about the truth of a sentence s when $\mathcal{M}$ believes that s is not necessarily true (i.e., $\Gamma \nvdash_{MSL} s$) or that s is false (i.e., $\Gamma \vdash_{MSL} \neg s$), and yet $\mathcal{M}$ believes $\mathcal{D}$ to be predisposed to believe that s is true (i.e., $\Sigma \vdash_{ToM} s$). To deceive $\mathcal{D}$ about the truth of s, $\mathcal{M}$ constructs an argument concluding s, and does so in a way that maximizes the argument's credibility (i.e., believability, persuasive force). $\mathcal{M}$'s process of argument construction is similar to that of proof-search in automated reasoning. $\mathcal{M}$ consummates the intended deception by articulating to $\mathcal{D}$ the most credible argument found for the truth of s. (Note that the most credible argument for s may well be normatively valid; this can occur when $\mathcal{M}$ believes that $\mathcal{D}$ already holds an erroneous belief, and s saliently follows, in part, from the erroneous belief in a perceptually, and logically, valid way. When $\mathcal{M}$ believes that $\mathcal{D}$'s beliefs are correct, $\mathcal{M}$'s argument for the truth of s is necessarily illusory.)

7 GENERATING ILLUSORY ARGUMENTS

There is no shortage of techniques available for generating illusory arguments. Here we quickly encapsulate three algorithms for doing so. For the first algorithm consider the following valid rule (R) of deductive inference:

$$\frac{(\phi \rightarrow \psi) \wedge (\neg\phi \rightarrow \psi)}{\psi}$$

Generally speaking, the validity of R can be grasped by a college-educated subject S. In light of this, an algorithm-sketch for generating a corresponding sophistic argument, assuming R (and the like) as input, is to first modify this rule to produce a variant R' that inherits the air of plausibility of the original, but which is nonetheless invalid; and to then generate, in natural language, an argument that conforms to R'. This is how the illusory English argument we presented above (§5) "showing" that there is an ace in the hand can be automatically generated. R' in the case of this "demonstration" is

$$\frac{(\phi \rightarrow \psi) \vee (\neg\phi \rightarrow \psi)}{\psi}$$

which is an invalid rule of inference.

For the second algorithm we use a set of inference rules (including some formal fallacies), and a rule-application procedure for enumerating sequences of inferences ordered by length. (Note that some applications of these rules may be perceived as invalid.) To produce an argument apparently demonstrating that ψ follows from Φ, we iterate over inferential sequences that conclude ψ from Φ, testing each for perceived validity. For those sequences perceived as valid, our measure of credibility is computed, and the most credible sequence is retained. The algorithm terminates when a threshold for credibility is surpassed, or when the time alloted for search is exhausted.

The third algorithm recasts argument construction as a graph-theoretic path-finding problem. Nodes in the graph represent states of a knowledge-base, edges characterize perceptually valid inferences,

and paths represent perceptually valid inferential sequences. By suitably defining path-length in terms of our valuation of credibility, constructing a most-credible argument that ψ follows from Φ reduces to finding a shortest path from the state where the knowledge-base contains only Φ to any state where the knowledge-base contains ψ.

8 PURELY PERSUASION

Our emphasis so far has been on persuading humans of falsehoods, i.e., the production of belief without knowledge via illusory arguments. While our approach to mechanical sophistry facilitates this type of deception, it is not limited to it. Our approach can equally persuade humans of truth, i.e., produce belief that is knowledge. This is so because the difference between a corruptive, illusory argument and an intuitive (i.e., persuasive) explanation of truth is simply whether (i) the argument's premises are true, and (ii) the reasoning articulated is logically valid. When our machine $\mathcal{M}$ is charged with arguing for a false conclusion, it attempts the corruption of another's mind (e.g., to instill erroneous belief).[8] But when $\mathcal{M}$ is charged with arguing for a true conclusion using only logically valid reasoning, it attempts the education or restitution of another's mind (e.g., to instill knowledge, to disabuse erroneous belief).[9] That is to say, by maximizing the credibility of valid reasoning, $\mathcal{M}$ tries to make truth comprehensible — to persuade humans of truths that they would otherwise likely not apprehend.

We do not aspire to a new philosophic theory of rhetoric. Instead, we are answering the charge that theoreticians and compositionists in Western rhetoric "have effectively halted the development of non-philosophic, sophistic theories of rhetoric and the realization of their practical possibilities" [16]. Our aim, ultimately, is an empirically-grounded, falsifiable theory of the relative persuasiveness of arguments to humans, and computational methods for generating persuasive arguments. The two combined are useful in the bending persuasion toward whatever end one desires.

ACKNOWLEDGEMENTS

We are indebted to Joshua Taylor, Michel Ingham, and three anonymous referees for insightful criticisms and comments (responses to which are woven into the present version of the paper).

REFERENCES

[1] D. Ariely, *Predictably Irrational*, Harper Collins, New York, NY, 2008.
[2] *Mental Logic*, eds., M. D. S. Braine and D. P. O'Brien, Lawrence Erlbaum, Mahwah, NJ, 1998.
[3] S. Bringsjord, E. Bringsjord, and R. Noel, 'In Defense of Logical Minds', in *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, eds., M. A. Gernsbacher and S. J. Derry, pp. 173–178, Mahwah, NJ, (1998). Lawrence Erlbaum.
[4] N. Chater and M. Oaksford, 'Rationality, rational analysis, and human reasoning', in *Psychology of Reasoning*, eds., K. Manktelow and M. C. Chung, 43–74, Psychology Press, New York, NY, (2004).
[5] C. I. Chesñevar, A. G. Maguitman, and R. P. Loui, 'Logical Models of Argument', *ACM Computing Surveys*, **32**(4), 337–383, (2000).

[8] Namely, $\mathcal{M}$ attempts one or more of the following (adapted from [7]): (i) contribute to another acquiring the belief in a falsehood; (ii) contribute to another continuing in the belief in a falsehood; (iii) contribute to another ceasing to believe in a truth; (iv) contribute to preventing another from acquiring the belief in a truth.

[9] Namely, $\mathcal{M}$ attempts one or more of the following: (i) contribute to another acquiring the belief in a truth; (ii) contribute to another continuing in the belief in a truth; (iii) contribute to another ceasing to believe in a falsehood; (iv) contribute to preventing another from acquiring the belief in a falsehood.

[6] R. M. Chisholm, *Theory of Knowledge*, Prentice Hall, Englewood Cliffs, NJ, 1966.
[7] R. M. Chisholm and T. D. Feehan, 'The Intent to Deceive', *Journal of Philosophy*, **74**(3), 143–159, (1977).
[8] B. de Finetti, 'La prévision: Ses lois logiques, ses sources subjectives', *Annales de l'Institute Henri Poincaré*, **7**, 1–68, (1937).
[9] F. Fitch, *Symbolic Logic*, Ronald Press, New York, NY, 1952.
[10] F. Grasso, A. Cawsey, and R. Jones, 'Dialectical argumentation to solve conflicts in advise giving: a case study in the promotion of healthy nutrition', *International Journal of Human-Computer Studies*, **53**(6), 1077–1115, (2000).
[11] P. N. Johnson-Laird, 'Rules and Illusions: A Critical Study of Rips's *The Psychology of Proof*', *Minds and Machines*, **7**(3), 387–407, (1997).
[12] P. N. Johnson-Laird, *How We Reason*, Oxford University Press, New York, NY, 2006.
[13] G. Lakemeyer, 'Steps Towards a First-Order Logic of Explicit and Implicit Belief', in *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning about Knowledge*, ed., J. Y. Halpern, pp. 325–340, San Francisco, CA, (1986). Morgan Kaufmann.
[14] H. J. Levesque, 'A Logic of Implicit and Explicit Belief', in *Proceedings of the Fourth National Conference on Artificial Intelligence*, pp. 198–202, Menlo Park, CA, (1984).
[15] M. Manzano, *Extensions of First Order Logic*, Cambridge University Press, Cambridge, UK, 1996.
[16] R. Marback, *Plato's Dream of Sophistry*, University of South Carolina Press, Columbia, SC, 1999.
[17] P. McBurney, S. Parsons, and M. Wooldridge, 'Desiderata for Agent Argumentation Protocols', in *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pp. 402–409, New York, NY, (2002). ACM Press.
[18] M. Piattelli-Palmarini, *Inevitable Illusions*, John Wiley & Sons, 1994.
[19] Plato, 'Gorgias', in *The Collected Dialogues of Plato*, eds., E. Hamilton and H. Cairns, 229–307, Princeton University Press, Princeton, NJ, (1971/380 BC). tr., W. D. Woodhead.
[20] *Cognitive Illusions*, ed., R. Pohl, Psychology Press, New York, NY, 2004.
[21] J. L. Pollock, *Cognitive Carpentry*, MIT Press, Cambridge, MA, 1995.
[22] H. Prakken, 'Formal systems for persuasion dialogue', *The Knowledge Engineering Review*, **21**(2), 163–188, (2006).
[23] H. Prakken and G. A. W. Vreeswijk, 'Logics for Defeasible Argumentation', in *Handbook of Philosophical Logic*, eds., D. M. Gabbay and F. Guenthner, volume 2, 219–318, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2nd edn., (2002).
[24] D. Premack and G. Woodruff, 'Does the chimpanzee have a theory of mind?', *Behavioral and Brain Sciences*, **1**(4), 515–526, (1978).
[25] F. P. Ramsey, *The foundations of mathematics and other logical essays*, Routledge & Kegan Paul, London, UK, 1931.
[26] C. Reed and F. Grasso, 'Recent Advances in Computational Models of Natural Argument', *International Journal of Intelligent Systems*, **22**(1), 1–15, (2007).
[27] E. Reiter, R. Robertson, and L. M. Osman, 'Lessons from a failure: Generating tailored smoking cessation letters', *Artificial Intelligence*, **144**(1–2), 41–58, (2003).
[28] K. J. Rinella, S. Bringsjord, and Y. Yang, 'Efficacious Logic Instruction: People are not Irremediably Poor Deductive Reasoners', in *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, eds., J. D. Moore and K. Stenning, pp. 851–856, Mahwah, NJ, (2001). Lawrence Erlbaum.
[29] L. J. Rips, *The Psychology of Proof*, MIT Press, Cambridge, MA, 1994.
[30] *The Mental Models Theory of Reasoning*, eds., W. Schaeken, A. Vandierendonck, W. Schroyens, and G. d'Ydewalle, Lawrence Erlbaum, Mahwah, NJ, 2006.
[31] S. E. Toulmin, *The Uses of Argument*, Cambridge University Press, Cambridge, UK, 1958.
[32] D. N. Walton and E. C. W. Krabbe, *Commitment in Dialogue*, State University of New York Press, Albany, NY, 1995.
[33] Y. Yang and S. Bringsjord, 'Mental Metalogic: A New Paradigm for Psychology of Reasoning', in *Proceedings of the Third International Conference of Cognitive Science*, eds., L. Chen and Y. Zhuo, pp. 199–204, Hefei, China, (2001). USTC Press.
[34] Y. Yang and S. Bringsjord, 'Mental metalogic and its empirical justifications', in *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, eds., R. Alterman and D. Kirsh, pp. 1275–1280, Boston, MA, (2003). Cognitive Science Society.

Persuasive gaze in political discourse

Isabella Poggi[1] **and Laura Vincze**[2]

[1] Dipartimento di Scienze dell'Educazione – Università Roma Tre; [2] PhD Student, Dipartimento di Linguistica – Università di Pisa
[1] mailto:poggi@uniroma3.it [2] mailto:l.vincze@ling.unipi.it

Abstract. The paper investigates the use of gaze in political communication. A study is presented on the electoral debates of two politicians in Italy and France. Fragments of their verbal discourse were analysed in terms of their hierarchy of goals, and their gaze communication was described and classified in terms of the persuasive strategies they pursue, *logos*, *ethos* or *pathos*. The results show that the pattern of gaze persuasive strategies is coherent with the hierarchy of goals pursued in the verbal discourse.

1 INTRODUCTION

The field of Persuasive technology includes the "design, research and analysis of interactive computing products created for the purpose of changing people's attitudes or behaviours" Fogg [1]. Persuasive systems are presently being designed for interaction in several domains, from health Mazzotta [2] to politics, and they use various kinds of interfaces, from those using Natural Language Generation only in the written verbal modality, to those that exploit Embodied Conversational Agents, that communicate through synthetic voice, and display gestures, head movements, body postures and gaze. But, what can be seen as persuasive in gaze behaviour? Are there some uses of gaze that are specifically aimed at persuading? If we think of some human behaviours which do, in fact, have a goal of influencing others, like, for instance, flirting or seduction, we can well say that some uses of gaze have a somehow "persuasive" import. However, for sure one could not use a seductive glance during a political talk show. In this paper, we explore the issue whether gaze can be used in a specific persuasive way in political discourse: after presenting a study that answered the question for gesture, and an overview of the semantic potentialities of gaze, we analyse gaze behaviour in an Italian and a French politician. Finding the features that give gaze a persuasive import can help to build Embodied Conversational Agents to be used in Persuasive technology by exploiting also gaze communication with a persuasive function.

2 RHETORICS AND BODY COMMUNICATION

The importance of bodily behaviour in persuasive discourse has been acknowledged back since the ancient Roman treatises of Rhetorics, like Cicero [3] and Quintilian [4]. In the rhetorical tradition, gestures and head movements have been studied as an indispensable part of "Actio" (discourse delivery), since they were credited with the capacity of fulfilling various communicative functions. By gestures and other body movements we can summon, promise, exhort, incite, approve, express apology or supplication, display emotions like regret, anger, indignation, adoration, depict or point at objects. For example, Quintilian [4] in his work provides detailed hints, mainly with a normative intent, about which movements may be more or less effective in portraying a particular image of the orator, which make him more similar to a comic actor, which can excite the audience or so.

In recent literature, some studies overview various aspects of the body's relevance in political communication Atkinson [5], like the use of pauses and intonation to quell the applause Bull [6], or facial expression and other bodily behaviours Frey [7], Bucy & Bradly [8]. Other studies that can give hints as to which nonverbal behaviours are ineffective or even prevent persuasion, are those concerning the detection of deceptive behaviours, like Ekman [9] and DePaulo et al. [10]. For example, self-manipulation, talking faster, averting eyes have a negative effect on persuasion in that they are felt as – and may be in fact – a cue to deception. Two recent studies directly concerned with the impact of gestural communication on political discourse are Calbris [11] and Streek [12]. The former analyses the gestures of Lionel Jospin as a way to understand the intimate expression of his political thought: for example, the metaphors exploited by his manual behaviour – whether he uses the left or right hand, and with which shape – can express abstract notions like effort, objective, decision, balance, priority, private or public stance; but they also fulfil discourse functions: they can delimit or stress, enumerate or explicate the topics of discourse.

Streeck [12] in analysing the gestural behaviour of the Democratic candidates during the political campaign of 2004 in USA, shows how important bodily behaviour may be in the political persuasion: as to the transmission of factual information he observes how the tempo of body movements and their relation to speech rhythm provide information about discourse structures, distinguishing background from foreground information; but as to the very persuasive effect he even attributes the defeat of Howard Dean to the frequency of his "finger wag", a "hierarchical act" that might have given an impression of presumption and contempt toward the audience. Also among the gestures analysed by Kendon [13] some are used with a persuasive intent: for example, the "ring" gestures, that bear a meaning of 'making precise' or 'clarifying', and are used every time this clarification is important "in gaining the agreement, the conviction or the understanding of the interlocutor" (p. 241).

If gesture is so important in conveying information that is effective in persuasion, also facial behaviour could be relevant in this connection. For example, in Italian political talk shows, while a politician is talking in a talk show, often the cameras record the facial expressions of his opponents, which are sometimes very communicative and may have a counter-persuasive role. Yet, not so much literature has been devoted to the persuasive impact of facial expression and gaze in persuasion.

3 COMMUNICATIVE GAZE

In this work we focus on the use of gaze in political discourse.
Eyes are used firstly for vision and then possibly to help thinking, for instance during memory retrieval or when thinking before speaking; but they are also used for communicating. Gaze as a communicative behaviour has been studied mainly since the seventies Kendon [14]; Argyle & Cook [15]; Duncan & Fiske [16], but it is presently an emerging topic of research, and its functions are studied in depth. As is clear from literature, gaze plays an important role in the management of face-to-face interaction, first in the very moment of deciding whether to engage in interaction, since it indicates the level of interest to interaction, then in establishing the focus of shared attention, and finally in managing the flow of conversation: it signals turn-taking manoeuvres Goodwin [17]; Rossano [18] provides rich and articulated feedback Allwood et al. [19], Heylen [20]; helps people's own communication management Allwood et al. [21]. Moreover, the importance of gaze is witnessed in mother-child interaction Trevarthen [22] and in classroom interaction Rosenthal & Jacobson [23]; Taeschner et al. [24] that are both sustained by joint attention and mutual gaze. Even, hypotheses about the physical structure of human's eyes, with a white sclera so different from non-human primates, have been put forward, as a way to allow eye detection and enhance shared attention Tomasello et al. [25].
Actually, this body of research is in general focused only on one aspect of gaze communication: eye direction. But this is not the only pertinent dimension of the gaze signal; many other aspects of gaze behaviour and of the whole eye region are relevant from a communicative point of view. This region can be distinguished into different sub-regions – eyebrows, eyelids, eye-sockets, wrinkles – and each of these parts has a role in conveying meanings; so much that an "optology" – a "phonology" of gaze – can be written down Poggi [26]; Poggi & Pelachaud [27] by finding out the parameters of gaze, in the same way as it has been done for the signs of sign language of the deaf and the symbolic gestures of the hearing.
Also on the meaning side, the function of gaze is not only to establish the setting for interaction, but to *tell things*: gaze does convey specific meanings, that can be discovered and listed in a systematic way.
Research also in these domains has been done; see for example the conversational and emotional signals of the eyebrows studied by Ekman [28]; the syntactic role of eye-gaze in Sign languages Baker & Padden [29]; the repertoire of gaze meanings of Kreidlin [30] founding an "Oculesics", and the analogous attempt of Poggi [26], Poggi [31] to write down a lexicon of gaze. All of these attempts have had a resonance in the construction of Embodied Agents, which have started to exhibit realistic communicative gaze behaviour Pelachaud & Poggi [32]; Maatman [33]; Heylen [20].
Notwithstanding this emerging research about gaze, its function in persuasive discourse has not been studied. Here we investigate the persuasive role of gaze in political discourse.
In the next sections we present a model of persuasion based on the notions of goal and belief and a research on the persuasive import of gesture in political discourse. Then we present an annotation scheme to describe and classify gaze in persuasive discourse, and we analyse some fragments of political discourse as to its gesture and gaze, aiming to find out different patterns of gaze communication in persuasive discourse.

4 PERSUASION AS A HIERARCHY OF GOALS

The model we adopt for the analysis of persuasion is one in terms of a goal and belief view of mind, social interaction and communication Castelfranchi & Parisi [34]; Conte & Castelfranchi [35]; Poggi [36]; Miceli et al. [37]. Persuasion is an act aimed at social influence, with social influence defined as the fact that an Agent A causes an increase or decrease in the likeliness for another Agent B to pursue some goal GA. In order to have B more likely pursue a goal GA, A must raise the value that GA may have for B, and does so through having B believe that pursuing GA is a means for B to achieve some other goal GB that B already has, and considers valuable. This definition encompasses different kinds of social influence, ranging from education to threat, promise, the use of strength and so on. Persuasion is a type of influence that 1) is pursued through communication, and 2) leaves B free of either pursuing the goal GA proposed by A or not. To persuade B to have GA as a goal of his, A must convince B, that is, induce B to believe with a high degree of certainty, that GA is worth pursuing since it is a sub-goal to some goal GB that B has. In order to do so, A can make use of three different strategies, as already stated by Aristotle [38] : *logos* (the logical arguments that support the desirability of GA and the link between GA and GB); *ethos* (A's personality, his intellectual and moral reliability); and *pathos* (the extent to which A, while mentioning the pursuit of goal GA, can evoke the possibility for B to feel pleasant emotions or to prevent unpleasant emotions).
In order to persuade other people we produce communicative acts by exploiting different modalities – we can use written texts, graphic advertisement, words, intonation, gestures, gaze, facial expression, posture, body movements: we thus make multimodal persuasive discourses, that is, complex communicative plans for achieving communicative goals. Any discourse can be analysed as a hierarchy of goals: a communicative plan in which each single communicative act (either verbal or non verbal) aims at a specific goal. And each goal may also aim at one or more super-goals in turn: further goals for which the first goal is a means. For example, if I say "Are you going home?" my literal goal is to ask you if you are going home, but through this I may aim at the super goal of asking for a lift. Two or more communicative acts may have a common super goal. For example, saying "I am here with this face" plus saying "this is the face of an honest person" may aim at the super goal of implying "I am an honest person". A discourse (both a unimodal and a multimodal one) is a sequence of communicative acts that all share a common super goal. For example, in a pre-election discourse, all the sentences, gestures, face and body movements aim at one and the same common super goal: "I ask you to vote for me". In a persuasive multimodal discourse, not only sentences, but also gestures, gaze, head movements and other signals may pursue, through their literal goals and their intermediate and final super goals, a *logos*, *ethos* or *pathos* strategy.

5 PERSUASIVE GESTURES

In a previous work, Poggi & Pelachaud [39] have investigated the impact of gestures in persuasive discourse. Some fragments of persuasive multimodal discourse were analysed as to their global meaning and their persuasive structure, and the gestures

performed during discourse were annotated. For each gesture, the signal side was analysed in terms of its parameters of handshape, orientation, location and movement, including its expressivity parameters Hartmann [40]: temporal extent, spatial extent, fluidity, power and repetition. It was found that there are not gestures whose meanings we can utterly define "persuasive"; rather, some gestures, or sometimes simply some parameters in their expressivity, convey some kinds of information that are typically contained in the cognitive structure of persuasive discourse. Types of information necessarily conveyed in persuasion are those somehow linked to the communicative strategies of *logos*, *ethos* and *pathos*: in particular, the following meanings.

1. *Importance*. If something is important, to obtain it will be a high value goal that you want to pursue. Gestures that convey the meaning "important" mention the high value of a proposed goal, to convince the Addressee to pursue it. This meaning is typically contained in gestures that convey performatives of incitation or request for attention, or other gestures like in [40]: *"the finger bunch"*, that convey a notion of importance as their very meaning; but "important" is also the meaning of *beats*, since every beat stresses a part of a sentence or discourse, communicating "this is the important part of the discourse I want you to pay attention to and to understand". Finally, this can be the meaning of irregularity or discontinuity in movement.
2. *Certainty*. To persuade you I must convince you, that is, cause you to have beliefs with a high degree of certainty, about what goals to pursue (their value, importance) and how to pursue them (means-end relations). To induce certainty in you, I may need to show self-confident and certain about what I am saying. This is why gestures that convey high certainty, like in Kendon [41]: *"ring"*, may be persuasive.
3. *Evaluation*. To express a positive evaluation of some object or event implies that it is a useful means to some goal; thus, to bring about that event or to obtain that object becomes desirable, a goal to be pursued. In a marketplace, to convince someone to buy a food, a "*cheek screw"* (rotating the tip of the index finger on cheek), that means "good", "tasty", made by a grocer, would be a good example of persuasive gesture.
4. *Sender's benevolence*. In persuasion not only the evaluation of the means to achieve goals is important, but also the evaluation of the Persuader: the Sender's *ethos*. If I am benevolent to you – I take care of your goals – you can trust me, so if I tell you a goal is worthwhile, you should pursue it. A gesture driven by the ethos strategy of showing one's moral reliability is the gesture, quite frequent in political communication, of *putting one's hand on one's breast* to mean "I am noble, I am fair": Serenari [42].
5. *Sender's competence*. Trust implies not only benevolence but also competence. If I am an expert in the field I am talking about, if I am intelligent, efficient, you might join with me and pursue the goals I propose. A politician, in talking of quite technical things concerning taxes, uses his right hand curve open, with palm to left, rotating rightward twice, meaning that he is passing over these technicalities, possibly difficult for the audience; but at the same time the relaxed appearance of his movement lets you infer that he is smart because he is talking of such difficult things easily, and unconstrained. This provides an image of competence.
6. *Emotion*. Emotions trigger goals. So A can express an emotion to affect B by contagion and thus induce him to do or not to do some action. In talking about his country an Italian politician, moving his forearm with short and jerky movements of high power and velocity, conveys the pride of being Italian in order to induce the goal of voting for him.
7. Among these types of information, Emotion (n.6) typically makes part of a pathos strategy; the Sender's benevolence and competence (n.5 and 4) and certainty (n.2), are clearly ethos information; while the elements of importance and evaluation (n. 1 and 3) are generally conveyed through a logos strategy. Nonetheless, these categories can merge with each other: for example, expressing an emotion about some possible action or goal may imply it is an important goal for me, and should be so for you. In this case, at a first level there is a pathos strategy – the goal of inducing an emotion, but this pathos is aimed at demonstrating the importance of the proposed goal, thus conveying a logos strategy at the indirect level.

6 PERSUASIVE GAZE

We present an observational study that investigated the persuasive impact of gaze in political discourse.

6.1 Hypothesis

Our working hypothesis, like in the study above, is that the persuasive import of gaze, just as that of words and gestures, depends on the meanings it conveys. Therefore, to assess how persuasive the gaze exhibited in a discourse can be, you have to assess its meanings.

But what are the meanings of gaze? According to the semantic typology proposed by Poggi [43], any communicative signal of any modality can provide one of three types of information concerning: 1. the World: concrete and abstract events and entities (objects, persons, animals, times and places); 2. the Sender's Identity: stable characteristics of the Sender (sex, age, culture, personality, image and self-image); 3. the Sender's Mind: his/her mental states (beliefs, goals, emotions) concerning the content and structure of the discourse s/he is delivering. Thus every gaze signal can be classified as to the type of information it bears. As to Information on the World, directing our eyes to a different direction from that pointed by our face, we may perform a deictic gaze, that is, point at an object or person; by squeezing eyes we may refer to something physically of conceptually "small", thus conveying a concrete or abstract property of something, for example, a "little" object or a "subtle" concept.

Other gaze items inform about the Sender's Identity (ethnic or cultural roots, personality): both the identity one cannot help showing – see the shape of the eyelids that distinguishes a Chinese from an Italian woman – and the identity one deliberately wants to project – see the *eyebrows raised with half-open eyelids* that show you are haughty Poggi & Roberto [44]. Finally, many gaze items are "Mind Markers" Poggi [45], signals conveying Information about the Sender's Mind: beliefs, goals and emotions referred to what we are talking about. Within Belief markers, we can inform about the degree of certainty of the beliefs we are mentioning: a *light frown* tells that I am serious, not kidding in saying what I say, while *raising eyebrows with not very open eyes* means I am a bit uncertain. Other signals

provide metacognitive information, i.e. they inform about the source of what we are saying (*looking downward leftward* tells the information we are going to provide is to be retrieved from our long-term memory); or else they inform of the mental state we are in (*staring out into space* or closing the eyelids means I am in concentration).
Within "Goal Markers", some express a performative, that is, the act a Sender has the goal to perform by his communication: a long lasting stare in the eyes of the Addressee is a defying gaze. An eyebrow raising marks the comment of our sentence, thus distinguishing what we want to stress vs. what we take for granted. Again, we define as metadiscursive gaze a light closing of the eyelids, which means I am going to skip a topic in my discourse; or a raising of the eyebrows that corresponds to saying "but", signalling a contrast between two beliefs. Finally, to manage turn-taking in conversation we gaze at the Speaker to take the floor; in providing backchannel we can frown to show we don't understand or we don't agree, or slightly close eyelids to say we are following the Interlocutor's discourse.
Finally, some gaze items, like the raised eyebrows and wide open eyes of surprise or the bright eyes of enthusiasm, are Emotion Markers, informing about the Sender's emotions.
Moreover, according to Poggi [45]; Poggi [43], all signals may have, beside their literal meaning, an indirect meaning, one that the Sender wants the Addressee to understand through automatic or context dependent inferences, and that can be very different from the literal one. Also a gesture or gaze can be classified in one of the types above both as to its literal meaning and to its indirect meaning, with the two classifications possibly being different. For example, *raising eyebrows with wide open eyes* showing surprise is an Emotion Marker at the literal level, but its indirect meaning can tell "I can't believe it", thus working as a backchannel Heylen [20].

6.2 Method

To assess the role of gaze in the persuasive structure of political discourse, we analysed some fragments of electoral debates held, respectively, by Romano Prodi, who won the elections in Italy in 2006, and by Ségolène Royal, who was defeated in France in 2007. For some fragments of both debates, the verbal communication was analysed in terms of their hierarchy of goals, aiming to find the structure of goals and supergoals and to detect the use of *logos*, *ethos* and *pathos* strategies. For other fragments, the Speaker's gaze behaviour was described and classified in an annotation scheme.

6.3 Analysis in terms of hierarchy of goals

Figure 1 shows the analysis of a fragment of one minute drawn from Ségolène Royal's debate. It was held in the studios of the French channel France 2, after the first round, when she came second with 25,87%, after Nicolas Sarkozy. In the political show "A vous de juger", Arlette Chabot interviews Mrs. Royal about her political vision and projects for France. Here is the fragment.

"Voilà, je n'ai aucune revanche à prendre, je n'ai aucune revendication, je n'ai pas d'enjeu personnel dans cette affaire, je ne suis liée à aucune puissance d'argent, je n'ai personne à placer, je ne suis prisonnière d'aucun dogme, et au même temps je sens que les Français ont envie d'un changement extrêmement profond. Et mon projet c'est eux, ce n'est pas moi, mon projet. Mon projet ce sont les Français et aujourd'hui le changement que j'incarne. Le changement, le vrai changement c'est moi"
(Look, I've got no revenge to take, I've got no claiming, no personal stake in this affair, I'm not bond to any financial power, I've got no one to place, I'm not prisoner of any dogma, and in the same time, I feel that the French people desire an extremely profound change. And my project it's them, my project it's not myself. My project is the French people and the change I embody today. The change , the real change, is me).

To analyse a fragment in terms of its hierarchy of goals means to figure out what are the communicative intentions of the Speaker. In this case, since we are analysing verbal behaviour, its hierarchy of goals is assumed as fairly aware, that is, the result of a somewhat conscious deliberation; in such a way that you could confirm your guess by asking the subject. Further, this is typically an analysis of the Speaker's intention, and does not take into account the actual effect of the persuasive efforts on the possible audience. When analysing a discourse as a hierarchy of goals, one has to put himself into the Speaker's mind, not into the mind of the Hearer.
According to Poggi's theory [36], we segmented the verbal discourse into its speech acts (written in italics), then we wrote down their literal goals and super-goals (i.e., the inferences each communicative act aims to induce, numbered as G1, G2 etc.), and we singled out the final goal of the fragment, by using arrows to represent their means-end relations.
In this fragment, Royal explains to the electors that she hasn't got any revenge to take (Speech Act SA1), any personal claim or benefit in this affair (2, 3, 5), nor is she bound to any financial power (4), thus implying (G7) that she does not work in her own interest; the only reason why she runs for President is for the sake of the French who wish for a change (SAs 7, 8, 9 , that aim at G8). By implying G7 and G8 she aims at demonstrating that she is altruistic (G6). At the same time, indirectly acknowledging (through SA 11) the good things done by her opponent (G10), she implies that she is fair (G5), with fairness and altruism bearing on an image of *benevolence* (G2), whose projection makes part of an *ethos* strategy. Meanwhile SAs 6 and 9, implying G9 ("I'm not prisoner of any dogma", and "I am the change") provide an image of flexibility, novelty, intelligence (G3): the *competence* side of the *ethos* strategy. Moreover, through SAs 10 and 12, she implies G11 ("Sarkozy is incapable to run France"), while through 13 and 14 she indirectly communicates G12 ("He is a coward"). So, symmetrically with the positive evaluations of competence and benevolence she implied about herself, she now provides two negative evaluations (G11 and G12) of her opponent both on the *competence* and on the *ethical* side. Further, that Sarkozy does not want to confront the balance of what he did (14) implies that the balance of what the political right side has done is negative (G13), so it is necessary to vote for the change (9), for the left (G4), for Royal (G1). The chaining of events and consequences from SAs 13 and 14 to G13 and from G13 to G9 and G4 may be seen as a *logos* strategy. Only the *pathos* strategy does not show so much in this fragment of discourse.
In her persuasive structure, Royal identifies the French people's desire for an extremely deep change (SA7) and she hooks her own goal of being elected to this goal, by communicating, through inference or explicit words (SAs 8 and 9) that *she* is the change: so voting for her (G1) is the means to their goal of bringing about change.

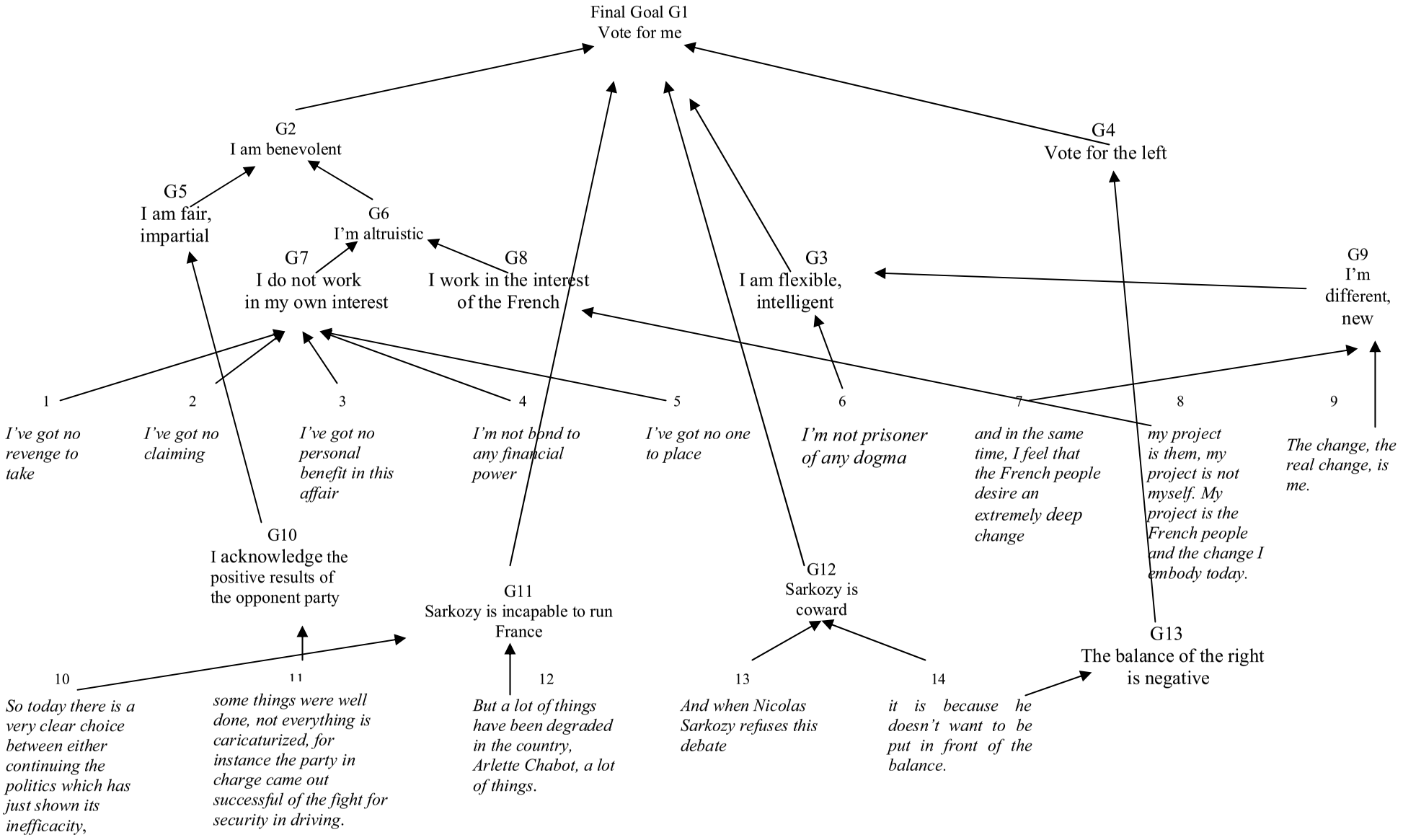

Figure 1. The persuasive strategy of a political discourse

6.4 Qualitative analysis of gaze

To analyse gaze in fragments of Royal's and Prodi's discourse we built an annotation scheme that describes and classifies each gaze item in terms of its signal, its meaning and its persuasive function (See Table 1). Specifically,

- Column 1 contains the time in the video;
- col. 2, the speech parallel to the gaze item under analysis;
- col. 3, a description of the gaze item in terms of its parameters Poggi & Roberto [43], like eyes direction (generally with respect to the direction of the head), actions of the eyebrows, position of the eyelids;
- col. 4, the literal meaning of the gaze item;
- col. 5: a classification of the meaning in col.4 according to the semantic typology above and possibly to its persuasive function, distinguished according to the persuasive strategy (*logos*, *pathos*, *ethos* benevolence, *ethos* competence);
- in cols. 6 and 7, the same analysis of cols.4 and 5 for possible indirect meanings of the gaze item.

Also in this case we write down the meanings we attribute to the Speaker's uses of gaze, on the basis of the lexicon of gaze hypothesised by Poggi [43]. They are the meanings and persuasive functions we think the Speaker has the goal (not necessarily a conscious goal) to convey, and we make no assumption as to whether they are in fact persuasive for the real audience. The difference from the meanings and functions we attribute to the Speaker as to the verbal fragment above, is that in gaze may be "meant" at a lower level of awareness.

Table 1 shows the analysis of two gaze items in Royal's discourse.

In example 1, while talking of the top managers who ruin the enterprises, like Mr. Forgeat (Col.2), Royal looks at the Interviewer, Arlette Chabot, with a fixed gaze (col.3) which means "I am serious, I do not let you avert gaze"(4): an information about Royal's personality, her being serious and determined, aimed at a strategy of *ethos competence* (col. 5), and possibly to indirectly conveying that she is one who struggles against injustice: again information on her personality, bearing on the moral side of *ethos*. Then Royal, leaning her head on the left, looks at the Interviewer obliquely and with half-closed eyelids, an expression of anger and indignation: information about her emotion, which she possibly wants to induce in the audience, thus pursuing a *pathos* strategy.

In 13, Royal refers to a proposal made by Sarkozy, that the unemployed should be induced to choose a job out of no more than two, and if they don't do so, they should lose their unemployment subsidy; she is arguing that this choice can only be acceptable if the conditions of the two jobs are not very punitive. So, while saying "you have to accept this job", she *looks down, first on the right then on the left*, as if looking at two things before deciding, thus referring to the choice between the two jobs. This is an iconic use of gaze, providing Information on the World, namely an action of choice, by miming it. After that, she *raises her eyebrows while keeping her eyelids in the défault position*: one more iconic gaze that means "order", miming the expression of someone who orders to the unemployed to make their choice. With these two gaze items Royal is playing the roles of both, the unemployed and the job proposer, thus dramatising the scene of Sarkozy's proposal. On the basis of the following argumentation, in which Royal is very critical about it, we can interpret her dramatisation as a parody, a way to make fun of Sarkozy's proposal.

In this way she conveys a negative evaluation of her opponent through a pathos strategy.

What this analysis shows is how coherent is the use of gaze with the whole of a Speaker's argumentation and persuasive plan.

n.	Speech	Gaze	Literal. M	Type	Indirect M.	Type
1 **48.10**	*Et aux hauts dirigeants qui abîment l'entreprise en faillite comme M. Forgeat* And what about the top managers who ruin the enterprises, like Mr. Forgeat,	Fixed gaze to Interviewer1 leaning head leftward, looks at her obliquely, with half-closed eyelids	I'm serious, I feel anger and indignation	ISI Personality ETHOS Competence ISM Emotion	I struggle against injustice I want you to feel indignation	ISI ETHOS Benevolence ISM PATHOS
13 **49.10**	*"Non, là, il faut… Il faut accepter cet emploi,* No, you have… you have to accept this job	She looks down, first right then left as if looking at two things to decide between them Eyebrows raised, Eyelids défault	Choice, choose a job I order you (to choose one)	IW Action ISM Performative	I am ridiculing Sarkozy's proposal → His proposal is too punitive	ISM Emotion PATHOS ISM Negative evaluation of opponent LOGOS

Table 1. Analysis of Royal's gaze items

6.5 Quantitative analysis

Out of the fragments analysed of Prodi's and Royal's discourse, we finally selected a sample of 20 items of gaze per each politician, and computed their persuasive functions. Table 2 and Figures 2 and 3 show the different distribution of gaze communication across the four different strategies: *logos, pathos, ethos competence* and *ethos benevolence.*

	Prodi	Royal
Length	53"	1'20"
Gaze items	20	20
Communicative units	25	25
Persuasive units	16	22
Logos	4	12
Pathos	2	3
Ethos competence	10	6
Ethos benevolence	0	1

Table 2. Prodi's and Royal's gaze

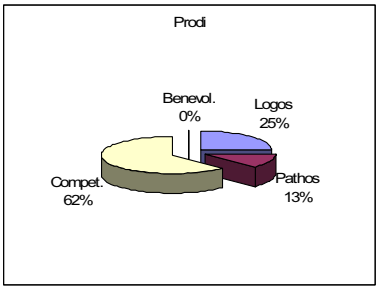

Figure 2. Persuasive strategies in Prodi's gaze

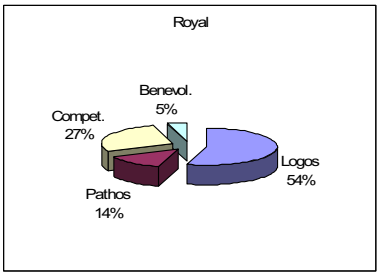

Figure 3. Persuasive strategies in Royal's gaze

6.6 Discussion

From the analysis of Prodi's and Royal's gaze, it results that the two differ as to the persuasive strategies of their gaze. Prodi's most frequent strategy (62%) is to show his competence to elicit the electors' trust, while he does not aim at showing his benevolence. He uses a logos strategy in 25% of his gaze, and pathos in 13%. Quite striking, instead, the frequency of logos strategy in Royal's gaze (54%) and then the appeal to her own competence (27%) which, however, does not rule out benevolence (5%). Finally, the pathos strategy is used slightly more than in Prodi's discourse (14% as opposed to 13%).

The high incidence of the *competence* and total lack of the *benevolence* strategy in Prodi's gaze is coherent with his need to enhance more his image of skill and political intelligence than that of an honest and altruistic person. Between Prodi and his opponent Berlusconi, the latter had the image of one who deals with politics more for the sake of his own financial interests than for Italians. On the other hand, the propaganda of the right would often ridicule Prodi by comparing him to a "mortadella" (a salted meat typical of his hometown, Bologna): a fat, cheap, not so tasty, not luxurious food, evoking someone who is over-soft, not very strong, skilled and determined. As for the high incidence of the *logos* strategy in Royal's gaze, two tentative accounts could be the goal to contrast stereotypes of feminine irrationality, or the French *esprit de géometrie*.

7 CONCLUSION

We have analysed some examples of the discourse and gaze of two politicians, and we have found out that the persuasive use of their gaze is coherent with the persuasive goals they pursue in their political communication.

This is only a first step in studying the persuasive impact of gaze in political discourse. A limitation of our work is that we did not study the effect of certain items of gaze, but only their goals. What we analysed here are the meanings that the Speaker has the (conscious or unaware) goal of communicating by his/her gaze items, and their persuasive import, without any consideration as to whether those meanings are effective in terms of real persuasive impact on the Addressee. Subsequent research could address this issue, by assessing if the meanings conveyed according to the Speaker's hierarchy of goals are actually persuasive for the audience. Moreover, through copy-synthesis or straightforward implementation, the face and gaze behaviour of humans – both the parameters that result in specific gaze items, and their corresponding meanings – could be reproduced in Embodied Conversational Agents. Evaluation studies could be carried on to assess both how humanlike and how effective the simulated is on the persuasion side, and their findings could help to build Embodied Conversational Agents that convey more persuasive messages.

REFERENCES

[1] B. J. Fogg. Persuasive Technology: Using Computers to Change What We Think and Do (Interactive Technologies). Amsterdam, Boston, Morgan Kaufmann Publishers, p.5. (2003).

[2] I. Mazzotta, F. de Rosis, V. Carofiglio. A user-adapted persuasion system in the healthy eating domain. "IEEE Intelligent Systems", 22, 6, pp.42-51 (2007)

[3] M.T. Cicero. Retorica ad Erennio. Patron, Bologna, Italy (1969).

[4] M. F. Quintilianus. Institutio Oratoria. Le Monnier, Firenze, Italy (1969)

[5] M. Atkinson. Our Master's Voices. The Language and Body Language of Politics. London, Methuen (1984).

[6] P. E. Bull. The Use of Hand Gestures in Political Speeches: Some Case Studies. In*: Journal of Language and Social Psychology*, 5: 102-118. (1986)

[7] S. Frey. Die Macht des Bildes. Der Einfluss der nonverbalen Kommunikation auf Kultur und Politik. Bern, Huber (1998)

[8] E. P. Bucy, S. D. Bradly. Presidential Expressions and Viewer Emotion: Counter Empathic Responses to Televised Leader Displays. *Social Science Information,* 43(1) : 59-94. (2004)

[9] P. Ekman. Telling lies. New York: Norton. (1985).

[10] B .M. De Paulo, J. J .Lindsay, B. E.Malone, L. Muhlenbruck, K. Charlton, & H. Cooper. Cues to deception. In *Psychological Bulletin*, 129, pp. 74-118. (2003).

[11] G. Calbris. L'expression Gestuelle de la Pensée d'un Homme Politique. Paris, Ed. du CNRS. (2003).

[12] J. Streeck. Gesture in Political Communication. A Case Study of the Democratic Presidential Candidates during the 2004 Primary Campaign. To be published in Research on Language and Social Interaction . Forthcoming

[13] A. Kendon. Gesture. Visible Action as Utterance. Cambridge. Cambridge University Press (2004)

[14] A. Kendon. Some Functions of Gaze Direction in Social Interaction. In : *Acta Psychologica*, 26, pp. 1-47. (1967)
[15] M. Argyle, M. Cook. Gaze and mutual gaze. Cambridge: Cambridge University Press. (1976)
[16] S. Jr. Duncan, D. H. Fiske. Face-to-Face Interaction: Research Methods, and Theory. Hillsdale: Erlbaum. (1977).
[17] C. Goodwin. Conversational organization. Interaction between speakers and hearers. New York: Academic Press. (2003)
[18] F. Rossano. When It's Over Is It really over?. On the Effects of Sustained Gaze vs. Gaze Withdrawal at Sequence Possible Completion. Paper at the 9th International Pragmatics Conference, Riva del Guarda, July 10-15, (2006).
[19] J. Allwood, J.Nivre, E. Ahlsén. On the Semantics and Pragmatics
[20] D. Heylen. A Closer Look at Gaze. Proceedings of the 4th International Joint Conference on Autonomous Agents and Multimodal Agent Systems 05. (2005).
[21] J. Allwood, E. Ahlsén, J. Lund., & Sundqvist. Multimodality in Own Communication Management. In J.Allwood, B.Dorriots & S. Nicholson (Eds.) *Multimodal communication. Proceedings from the Second Nordic Conference on Multimodal communication*, Gothenburg Papers in Theoretical Linguistics, Department of Linguistics, Goteborg University, pp. 43-62. (2005)
[22] Trevarthen, C. Emotions in Infancy: Regulators of Contact and Relationships with Persons. In K.R.Scherer & P.Ekman (Eds.), *Approaches to Emotion.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc., pp. 129-157. (1984).
[23] R. Rosenthal, L. Jacobson. Pygmalion in the Classroom. Holt, Rinehart and Winston, New York, USA (1968).
[24] T. Taeschner, P. Testa, M. Cacioppo. L'insegnante Magica. In T.Taeschner (Ed.), L'insegnante Magica. Roma: Borla. (2002)
[25] M. Tomasello, B. Hare, H. Lehnmann, J. Call. Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. In: *Journal of Human Evolution* 52, 3, pp. 314-320. (2007)
[26] I. Poggi. The Lexicon and the Alphabet of Gesture, Gaze, and Touch. In A.de Antonio, R.Aylett and D.Ballin (Eds.) *Intelligent Virtual Agents.* Berlin: Springer, p.235-236. (2001)
[27] I. Poggi, C. Pelachaud. Signals and Meanings of Gaze in Animated Faces. In P.McKevitt, S. O' Nuallàin, C.Mulvihill (Eds.), *Language,Vision, and Music. Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway*. 1996. Amsterdam. John Benjamins, pp.133-144, (2002).
[28] P. Ekman. About brows: Emotional and Conversational Signals. In M. von Cranach, K.Foppa, W. Lepenies, & D.Ploog (Eds.), *Human Ethology: Claims and Limits of a New Discipline: Contributions to the Colloquium.* Cambridge: Cambridge University Press, pp.169-248. (1979).
[29] C. L. Baker, C. Padden. Focusing on the Nonmanual Components of ASL In P.Siple (Ed.), *Understanding Language through Sign Languange Research*. New York: Academic Press. (1977).
[30] G. E. Kreidlin. *Neverbal´naia Semiotika: Iazyk tela i estestvennyi iazyk*. Moskva: Novoe literaturnoe obozrenie (2002).
[31] I. Poggi. Le parole del corpo. Introduzione alla comunicazione multimodale. Roma: Carocci (2006)
[32] C. Pelachaud, I. Poggi. Subtleties of Facial Expressions in Embodied Agents. In:. *Journal of Visualization and Computer Animation*, 31, 301-312. 2002).
[33] R. Maatman, I. Gratch, S. Marsella. Natural Behaviour of a Listening Agent. In Proceedings of the *5th International Conference on Interactive Virtual Agents*. Kos, Greece., pp. 24-31, (2005).
[34] C. Castelfranchi, D. Parisi. Linguaggio, Conoscenze e Scopi. Bologna: Il Mulino (1980).
[35] R. Conte, C. Castelfranchi. Cognitive and Social Action. London: University College (1995).
[36] I. Poggi. The Goals of Persuasion. In: *Pragmatics and Cognition* 13, 2005, pp.298-335 (2005).
[37] M. Miceli, I. Poggi, F. de Rosis. Emotional and Non-Emotional Persuasion *IJIS, Special Issue on Natural Argumentation* (2006).
[38] Aristotle. Retorica. Bari: Laterza (1973).
[39] I. Poggi, C. Pelachaud. Persuasive gestures and the expressivity of ECAs. In I.Wachsmuth, M.Lenzen, G.Knoblich (eds) *Embodied Communication in Humans and Machines*. Oxford, Oxford University Press.
[40] B. Hartmann, M. Mancini, C. Pelachaud. Formational Parameters and Adaptive Prototype Instantiation for MPEG-4 Compliant Gesture Synthesis. *Computer Animation 2002*, pp. 111-119 (2002).
[41] A. Kendon. Contrasts in Gesticulation. A Neapolitan and a British Speaker Compared. In C.Müller and R.Posner. (Eds.), *The Semantics and Pragmatics of Everyday Gestures*. Berlin: Weidler (2004).
[42] M. Serenari. Examples from the Berlin Dictionary of Everyday Gestures. In M.Rector, I.Poggi & N.Trigo (Eds.) *Gestures. Meaning and Use*. Porto: Edicoes Universidade Fernando Pessoa. (2003).
[43] I. Poggi. Mind, Hands, Face and Body. A Goal and Belief View of Multimodal Communication. Berlin: Weidler. (2007).
[44] I. Poggi, E. Roberto. The eyes and the eyelids. A compositional view about the meanings of Gaze. In E. Ahlsén P. J. Henrichsen, R. Hirsch, J. Nivre, A .Abelin, S. Stroemqvist, & S. Nicholson (Eds.), *Communicaion – Action – Meaning. A festschrift to Jens Allwood*. Dpt. Linguistics, Goteborg University, pp. 333-350. (2007)
[45] I. Poggi. Mind Markers. In M. Rector, I. Poggi & N. Trigo (Eds.), *Gestures. Meaning and use*. Porto: Universidade Fernando Pessoa, pp. 119-132. (2003).

Persuasive technology for shaping social beliefs of rural women: *Development of group based health information kiosk*

Vikram Parmar
Industrial Design Department,
Technical University of Delft,
The Netherlands
v.s.parmar@tudelft.nl

Abstract. This paper presents, the Group Based Information Kiosk (GIK), which was designed to influence health behaviours of rural women. The objective of the kiosk is to offer health information to rural women to increase their awareness about menses and maternal health. The design and development process of a GIK followed social cues of persuasive technology to increase perceived behaviour control of rural women. In order to measure user's level of engagement, a comparative study between the GIK and conventional health information system was conducted. The results of the interactive sessions with women of different age group and literacy level showed that, the GIK motivated women to challenge existing social beliefs and practices, thereby motivating them to follow correct health practices. In this paper, design process of GIK, preliminary results of the initial study, and future research plans are discussed.

1 INTRODUCTION

The delivery of relevant personal health information to improve the well being of rural population is still recognized as a large challenge for government and private sectors in rural India [9, 10, 11]. Despite several outreach programmes run by the government, non government organisations, and private sectors, there is still an alarming rate of deaths in rural area due to lack of primary health information [6, 8].

Recent studies have shown that one of the factors restricting the information exchange within the rural populace can be attributed to the social set-ups and orthodox social beliefs related to healthcare [4]. A study by Spector (1995) indicates that ignorance of these culturally divergent beliefs and traditional health care practices may lead to failure of health information system. However the existing health care approaches using both traditional and ICT medias do not consider the social beliefs while disseminating health care information. Additionally, the ICT based information systems disseminate general health care information and do not address domain specific information needs. Due to high percentage of illiteracy in rural areas user are highly dependant on community health worker to access health information.

Currently, persuasive technology has strongly emerged as a strategy for changing people's social attitudes [5]. Application of persuasive technology varies from persuading users to reduce energy consumption [2], assisting patients to remember their pills [13] and persuading young girls to avoid early pregnancy [5]. However, there is little evidence on how persuasive technology could be used in rural areas of developing countries, to addresses socio-cultural issues related to healthcare.

This paper aims to explore persuasive technologies that can be employed to improve the information content and interaction to disseminate health care information among rural populace. In our previous research [14], we identified social beliefs and practices related to maternal health and menstrual cycle among women. Based on these findings we developed a group based health information kiosk for rural women to persuade them to change their existing health practices. The kiosk includes audio visual aid to facilitate interaction between the literate and semi illiterate rural women. We incorporated the social beliefs and practices related to the health care issues, as stated by [14] in designing the information content of the kiosk. Additionally, we developed a persuasive environment by incorporating the group behaviour of rural woman in designing the system interaction.

We conducted a study to compare the level of engagement of the rural woman with the proposed group based systems to the conventional health information The comparison was done between two systems in terms of revisits, and users' capacity to challenge the social beliefs after the interactive session. Conventional kiosks offer general health information via a PC computer. The results of the study indicate that women interacting with the Group Based Information Kiosk (GIK) were persuaded positively about their incorrect health practices and were encouraged to challenge the existing social beliefs. The following sections describe the persuasive strategies applied in design of (GIK), and finally we conclude by discussing relevance of persuasive technology in planning information content for the rural context.

2 GROUP BASED INFORMATION KIOSK

Group Based Information Kiosk (see figure 1) is designed to offer personal health information related to maternal health and menses to rural women. This was developed to encourage them to challenge the existing social beliefs and to persuade them to change their incorrect health practices. The PC based system set up of GIK includes a monitor, a customised icon based keyboard

with 9 keys and a trackball. The information is presented by audio video aid in local language using PowerPoint, which can be accessed by the iconic keyboard using the 9 keys. PowerPoint platform was selected for information presentation due to its flexible and simple interface [7]. Power point was also selected because the content had to be easily updatable by the GIK operators, who were computer literate.

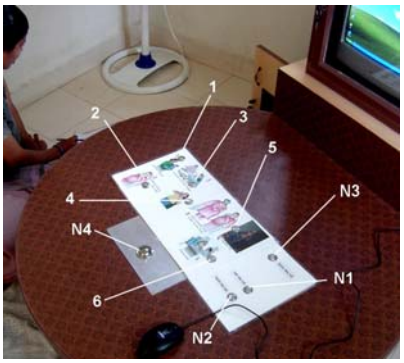

Figure-1 depicts design platform and layout

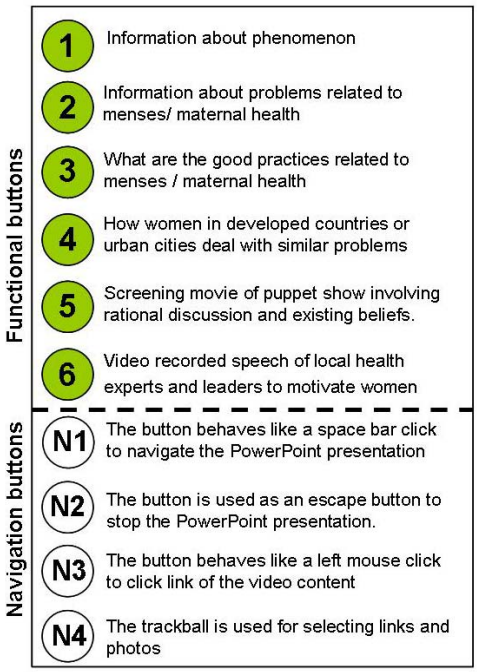

Figure-2 depicts explanation of each key

The healthcare content in GIK was designed by incorporating the social beliefs related to maternal health and menses found by [14]. The social beliefs related to maternal health included issues such as child delivery should only be done at home, pregnant mother should be given less food and pregnant ladies should not take any medical injections. The social beliefs related to menses included issues such as, only cotton clothes should be used during menses, during menses girl should not be allowed in the kitchen and menses is a punishment from God.

The GIK content is divided into six categories as depicted in (figure -2). Next to each input key, an icon is given to assist rural women in understanding the function of the concerned key. The system has flexibility to offer information sharing environment at individual or group level. Three local girls with knowledge of Microsoft office has been appointed as operators. Their responsibility is to assist rural women in interactive sessions, and update the kiosk content. The GIK is physically located in the community hall at a pilot village site in western India.

3 PERSUASIVE STRATEGIES IN GIK

(Table-1) depicts six information categories and corresponding persuasive strategies suggested by [5] that are employed to design the information content and physical form.

GIK components **Content** ***Input keys: Information detail***	**Persuasion strategies applied on the GIK**
Key-1: information about menses and maternal health.	Psychological clues*, Language*, Story telling technique, Audio-visual feedback in presentation.
Key-2: information about problems related to menses and maternal health.	Graphics depicting worst health condition of a women during maternal health, Story telling technique, Audio-visual feedback in presentation, Language, Psychological clues
Key-3: what are the good practices related to menses and maternal health.	Story telling technique, Audio-visual feedback in presentation, Language, Psychological clues
Key-4: how women in developed countries or urban areas deal with similar problems	Story telling technique, Audio-visual feedback in presentation, Language, Psychological clues
Key-5: screening health related videos	Social roles*, Using traditional folk songs, puppet shows, Story telling technique, Audio-visual feedback in presentation, Language, Psychological clues
Key-6: video recorded speech of local health expert and leader to motivate women.	Social roles, Story telling technique, Audio-visual feedback in presentation
Physical form	
Physical environment	Physical*, Social dynamics*, Social role
GIK form design	Social dynamics

Table-1 depicts six information categories and persuasive strategies applied in designing content and physical form

Physical clues*:* According to [5], computing technology can convey social presence through physical characteristic. Furthermore, physical attractiveness has a significant impact on social presence. In GIK, due to low height, rural women an easily sit on the floor to interact with the system. The sitting

position is culturally acceptable among rural women because it's a standard interaction posture among rural women, thus persuading women to sit for long hrs.

***Psychological clues**:* According to [5], computing product can lead people to infer, often subconsciously, that the product has emotions, preferences, motivations, and personality. In GIK, the following three animations of local personas were used to deliver the content: a) teenage girl for discussing menses issues, b) married woman for discussing maternal issues, and c) a doctor for answering their queries. Additionally, folk music, and puppet shows were used to denote the stories weaved around the personas. In all the 6 categories, the three personas presented existing beliefs and discussed its effects rationally. Due to similarity between the screen character and rural women, women could emotionally relate to beliefs and problems being discussed. As a consequence, women were persuaded to discuss beliefs among them selves.

***Language**:* According to [5], computing products can also use written or spoken language to convey social presence and to persuade. In GIK, during interactive sessions, each section of information item concluded with a provocative message, which persuaded rural women to press next input key and access additional information. A critical issue of varied literacy level was addressed by offering health information in local regional language.

***Social dynamics** :* According to [5], most cultures have set patterns for how people interact with each other such as, rituals for meeting people, taking turns, or forming lines. These rituals indicate the social dynamics. In GIK design process, current pattern of accessing information in public or private places, rituals of meeting people and community sensitive needs of rural areas were considered. GIK physical design offers flexibility to have individual or group based interaction with a rural women. Inviting women in groups includes family members such as sister in law, mother in law, and close friends of the rural women. This leads to interesting discussions in the sessions due to social dynamics.

***Social role**:* According to [5], human play authority roles, computers can also act in these roles, and when they do, they gain the automatic influence that comes with being in a position of authority. In rural areas, local doctors and village leaders are seen as authority and influential people. The word from these authorities has high value among rural women. In GIK, recorded videos speech of these authorities has been shown to persuade rural women towards following healthy practices during menses and maternal health.

4 INITIAL STUDY

A study was conducted to compare the GIK with the conventional health system to measure their level of engagement with the rural woman in terms of revisits to the system, and encouragement received to challenge the social beliefs. The study was conducted at the deployment site of GIK in villages in western India in two villages. In village A, the data was shown to (N=50) semiliterate and literate rural women on a conventional health information system. The content of the system was based on the available information related to menses and maternal health in the local hospital. In village B, the data was shown to (N=50) semiliterate and literate rural women through the proposed GIK. The content of the GIK was based upon existing social beliefs; furthermore, GIK offered flexibility to have interactive session in groups. The sessions were monitored by the researcher and the appointed local girls. In total, 100 rural women participated in the study. Participants were from 12-60 age groups. Both the villages were observed for two months. Local doctors were responsible for content verification and providing answers to participant queries related to menses and maternal health.

Figure-3 depicts Village-A participant in interactive session

Figure-4 depicts Village-B participants in interactive session

5 RESULTS

The results show that the Village B participants who interacted with the GIK were twice as much engaged in confronting the social issues as compared to participants from village A . Participant's engagement with the system was measured by two indicators, first amount and type of questions asked in the session, and second, number of re-visits to the health system after the first session. In total 23 questions from village-A and 58 questions from village-B was received from the interactive sessions. Received questions from the participants were categorised using affinity technique [3]. The following 6 categories were identified as prominent: questions related to practices, beliefs, challenging existing beliefs, data-addition, personal problems requiring privacy, and non-health related question. Based on frequency analysis, participant response on each category is depicted in table-2. and number of re-visits to their respected health system is depicted in table-3.

Questions asked	Village- A N= 50 women N (%)	Village- B N=50 women N (%)
1. Existing practices	5(10)	12(24)
2. Existing beliefs	3(6)	11(22)
3.Challenging existing beliefs	2(4)	7(14)

4. Data addition	8(16)	15(30)
5. Personal questions	2(4)	9(18)
6.Non health related questions	3(6)	4(8)

Table-2 provides details of the questions asked by the Village-A 23(100) and Village-B 58(100)

Number of re-visits	Village- A	Village- B
	N= 50 women N (%)	N=50 women N (%)
Participants	17 (34)	36 (72)

Table-3 provides details of the participants who re-visted the given health system after the first session

1. According to this study, participants from Village- B were twice as much engaged in confronting social issues than participants from Village- A. This could be attributed to the content design based on social beliefs and group based participation which led to interactive sessions in Village-B. This created peer pressure among similar background participants, also known as normative influence [5], which motivated participants to ask questions during the sessions.
2. Questions reported by Village-B regarding practices and challenging existing beliefs were slightly over twice than the response received from Village-A. In village-B, issues about existing beliefs and practices were openly discussed by the participants during the session. Many participants discussed their personal beliefs and practices to confirm their knowledge about maternal health.
3. Questions involving personal health and privacy were reported high in Village-B, 9(18%) sessions than Village-A, 2(4%). This implies that, issues which were considered private came out during group discussion. Group based sessions enabled women to share personal health issues in public.
4. In Village-B, 36 (72%) participants re-visited the GIK after their first session. Whereas, only 17(34%) participant re-visited the conventional health system from Village-A. The qualitative findings show that, participants from Village-A found information items relevant, but were not able to discuss with women of similar background during the sessions. Many participants also reported receiving too much information in one session and lack of discussion has led to less engagement. Therefore they are not keen on coming back. The strategy of inviting women in group in Village-B resulted in higher motivation and discussions between the participants.

6 CONCLUSIONS

The paper presented an exploration of persuasive technologies employed to improve the information content and interaction to dissemination of health care information among rural populace. Based on previously identified social beliefs and practices related to maternal health and menstrual cycle among women, a group based health information kiosk for rural women was developed. The results of the comparison study between the group based systems vs. a convention one indicate the benefits of deploying persuasive technologies to assist the rural women to confront their existing health practices. However, understanding and shaping beliefs doesn't guarantee change in health behaviour, unless intention of user to perform certain behaviour is not actually implemented [1]. In the next stage, longitudinal studies are planned to observe the changes in the user behaviour in relation to issues concerning maternal health and menses.

REFERENCES

[1] Ajzen, I. and A. S. R. Manstead (2007). Changing health-related behaviors: An approach based on the theory of planned behavior. New York, Psychology Press.

[2] Bang, M., C. Torstensson and C. Katzeff (2006). The PowerHouse: A Persuasive Computer Game Designed to Raise Awareness of Domestic Energy Consumption. Persuasive 2006, Springer.

[3] Beyer, H. and K. Holtzblatt (1998). Contextual Design: Defining Customer-Centered Systems. San Francisco, Morgan Kaufmann

[4] Bharat, S. (2001). India: HIV & AIDS-related discrimination, stigmatization and denial.

[5] Fogg, B. J. (2003). Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco, Morgan Kaufmann Publishers.

[6] Khanna, A., R. S. Goyal and R. Bhavsar (2005). "Menstrual practices and reproductive problems: A study of adolescent girls in Rajasthan." Journal of health management 7.

[7] Lin, L.-C. and W. O. Lee (2007). UI Toolkit for Non-Designers in the Enterprise Applications Industry. CHI 2007, San Jose, California, USA.

[8] Pakenham-Walsh, N. (2002). Strengthening Local Capacities to Create and Adapt Healthcare Information The Hague, International Institute of communication and Development.

[9] Pandey, P., A. Sehgal, M. Riboud, D. Levine and M. Goyal (2007). "Informing Resource- Poor Populations and the Delivery of Entitled Health and Social Services in Rural India." American Medical Association 298(16): 1867-1875.

[10] Prahalad (2006). "To create an impossibly low-cost, high-quality new business model, start by cultivating constraints." Strategy + Business(44).

[11] Sen, A. (1999). Development as Freedom. Oxford, Oxford University Press.

[12] Spector (1995). "Cultural aspects of women's health and health-promoting behaviors." Journal of Obstetrics and Gynecologic and Neonatal Nursing 24(3): 241-245.

[13] Sterns, A. A. and C. B. Mayhorn (2006). Persuasive Pillboxes: Improving Medication Adherence with Personal Digital Assistants. Persuasive 2006, Springer.

[14] Parmar, V., Keyson, D. V., Bont, CJPM, de (2007). Shaping social beliefs: A community sensitive health information system for rural India. In IFIP International Federation for Information Processing, Volume 241, Home Informatics and Telematics: ICT for the Next Billion, eds. Venkatesh,A., Gonsalves, T,. Monk, A,. Buckner, K,.(Boston: Springer), pp. 133-143.