Proceedings of the Symposium

Persuasive Technology and Digital Behaviour Intervention Symposium

A symposium at the AISB 2009 Convention (6-9 April 2009)
Heriot-Watt University, Edinburgh, Scotland

Symposium Chairs
Judith Masthoff
Floriana Grasso

Published by SSAISB:
The Society for the Study of Artificial Intelligence
and the Simulation of Behaviour
http://www.aisb.org.uk/

ISBN - 1902956745

Persuasive Technology and Digital Behaviour Intervention Symposium

A two-day symposium at AISB 2009 (6-9 April 2009).
http://www.csd.abdn.ac.uk/~jmasthof/Persuasive09/index.htm

PROGRAMME CHAIRS

Judith Masthoff, University of Aberdeen, UK
Floriana Grasso, University of Liverpool, UK

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? This symposium focuses on how digital technology can motivate and influence people. It brings together researchers, designers, and developers interested in computers designed to change attitudes and behaviours 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 Persuasive conference series (in Eindhoven, the Netherlands, 2006; Stanford, US, 2007; Oulu, Finland, 2008; Claremont, US, 2009), as well as the successful series of workshops on Computational Models of Natural Argument (an area overlapping with persuasion). This symposium covers the areas of persuasion systems, behaviour intervention technology and argumentation. It follows on from the successful Persuasive Technology Symposium held at the previous AISB. In 2008, we brought together researchers from distinct subfields of Computing Science (namely persuasive technology and argumentation). Now, we would like to extend this further to include Psychologists. Initial contact with this community has been established at the "Designing digital interventions to help overcome addictive behaviours" workshop in Windsor in 2008.

TOPICS

Topics of interest include but are not limited to:

- Behaviour intervention methods
- 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
  - Positive mood induction
  - 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 and behaviour intervention
- Ethics of persuasive technology
- Applications of persuasive technology and behaviour intervention, like in healthcare, education, e-commerce, politics

PROGRAMME COMMITTEE

Katie Atkinson, University of Liverpool, UK
Paul Beatty, University of Manchester, UK
Timothy Bickmore, Northeastern University, US
Guiseppe Carenini, University of British Colombia, Canada
Susan Ferebee, University of Phoenix, US
Nancy Green, University of North Carolina Greensboro, US
Marco Guerini, ITC-IRST, Povo-Trento, Italy
Helmut Horacek, University of the Saarland, Saarbrücken, Germany
Cees Midden, Eindhoven University of Technology, Netherlands
Hien Nguyen, University of Aberdeen, UK
Chris Reed, University of Dundee, UK
Fabio Paglieri, ISTC-CNR, Rome, Italy
Patrick Saint-Dizier, IRIT-CNRS, Toulouse
Falko Sniehotta, University of Aberdeen, UK
Oliviero Stock, ITC-IRST, Italy
Peter de Vries, Twente University, Netherlands
Doug Walton, University of Windsor, Ontario

Table of Contents

Measure Of Belief Change as an Evaluation of Persuasion

Pierre Andrews and **Suresh Manandhar** [1]

Abstract. In the field of natural argumentation and computer persuasion, there has not been any clear definition of the persuasiveness of a system trying to influence the user. In this paper, we describe a general evaluation task that can be instantiated on a number of domains to evaluate the beliefs change of participants. Through the use of a ranking task, we can measure the participant's change of beliefs related to a behaviour or an attitude. This general metric allows a better comparison of state of the art persuasive systems.

1 Motivation and Related Work

In novel fields of research, researchers often want to compare their approaches and the efficiency of their research. Thus, alongside the new field a research movement is created to develop robust evaluation frameworks that can provide comparative results and fair evaluations of research output for the field. For example, in the Information Retrieval field, the researchers have long studied different techniques of evaluation and selected the precision/recall measures, thus creating a framework of measures that can be used by all researchers and create evaluation campaigns such as the Text REtrieval Conferences (TREC `http://www.trec.nist.gov`).

The field of automated persuasion is attracting a growing interest in the research community, with new conferences and workshops every year [16, 19]. However, there has yet not been an agreed method for evaluating and comparing persuasive systems' output.

Existing research already provides examples of evaluation techniques for persuasion. For instance [4] uses a long term evaluation procedure to follow the change of students' behaviour when trying to persuade them to walk more. The measure of persuasiveness introduced by the authors is computed from the evolution of steps count for each participant, showing the change in walking behaviour of the students over one month. In this experimental setup, the researchers need a large amount of resources and time to provide pedometers to students, motivate them to use the system on a long term basis and wait for results; this amount of resources are not always available to all researchers. In addition, following a long term behaviour change is not an atomic setup and it is difficult to control for every external factors that can influence the user's behaviour.

[21] describes a smoking cessation program that tries to persuade participants to stop smoking through tailored letters. The users are asked if they think they will stop smoking in the month or six months following the reading of the letter. Participants were also asked if they had actually quit six month after the intervention. In this experiment, the authors show that there is no difference in the change of behaviour between the control group and the group that reads the tailored letters. The authors acknowledge that their experiment and trial was too small to show any statistical evidence. It is in fact difficult with such binary observation to extract enough data and it is a general problem to be able to find enough participants to follow on such a long term experiment.

In behavioural medicine, many measures have been developped to evaluate the changes in different mental constructs associated to behaviour change. [11], for example, proposes a questionnaire to evaluate the stage of change (see [20]) of participants within the pain treatment domain, while [15] develops a scale to measure the evolution of self efficacy in the domain of arthritis treatment.

Other persuasive system researches take a more concise approach by evaluating a change during the persuasive session of an external representation of the user's intentions towards a behaviour. For instance [8] evaluates an embodied conversational agent simulating a real estate agent by comparing the amount of money that clients are prepared to spend on a house before and after the interaction with the estate agent. The estate agent tries to convince users to buy a house fifty percent more expensive than what they are actually ready to spend. The persuasiveness of the system is evaluated by looking at the increase in the user's budget after the interaction. The measure is between zero percent to 100% increase relative to the target increase chosen by the system.

[18] tries to evaluate the effect of distance over persuasion for computer mediated communication. The author uses a setup following the desert survival scenario [14] where participants have to rank a set of items relating to their survival in the desert. After having given an initial ranking, the participants are then faced with persuasive messages relevant to these items and finally give a ranking of the same items after the persuasive session. The author uses as a measure of persuasion the distance between the participant's final ranking and the ranking of the persuader. [7] introduces a variation of the ranking task in the domain of house buying; instead of having to rerank a full list of items (houses in this case), the participants are persuaded to insert a new item in their initial ranking. This evaluation measures how many users actually chose the new alternative and where they ranked it in the initial ranking. These measures allow the authors to evaluate the persuasion and the effectiveness of the tailoring of the arguments.

We believe that a ranking task such as the one used by [18] can apply to different domains and be used as a common evaluation metric to compare persuasive systems. In this paper, we ground the validity of this ranking task in theory of persuasion and describe a formalisation of the ranking task that provides an evaluation metric for controlled experiments that can be more robust to external factor. It also provides a standard measure available in many domains and that can be compared between researches. We also conclude that there is a need for more research in persuasion evaluation frameworks to help the development of the automatic persuasion and natural argumentation field.

[1] University of York, United Kingdom, email: pierre.andrews@gmail.com; suresh@cs.york.ac.uk

2 Behaviour and Belief Change

When thinking of persuasion, the immediate indicator of success is a change in the behaviour of the persuaded subject; in fact, [17] proposed the following definition of *Persuasive Communication*:

> "Any message that is intended to shape, reinforce or change the responses of another or others." *from [22], p. 4*

It is generally accepted that the *"changed responses"* refers to either a change in behaviour or a change in the attitude towards the behaviour (see [22]). However, behaviours can take many forms and the method of evaluating of a change in behaviour will be different for every application domain. For instance [4] tries to evaluate a change in walking behaviour and uses the number of steps a user performs as a measure of behaviour change. In another health advice domain, [21] tries to convince participants to stop smoking, the evaluation output is thus the number of participants that stopped smoking. [4] describes a continuous evaluation value for each participant that is hard to port to other domains whereas [21] describes a binary value that does not provide powerful data for analysis but is easy to understand. Both evaluation methods consider a change of behaviour and provide the authors with a tool to demonstrate the persuasiveness of their system. However, it is difficult for the reader to make a comparison between the approaches' performances.

However, research in sociology and persuasive communication shows that intentions towards a behaviour can be modelled as a function of the user's beliefs about such behaviour and the social norms influencing the user. For instance, [1] presents the Theory of Reasoned Action that is designed to predict one's *intention* (I_B) to perform a particular *behaviour* (B) as a function f of one's *attitude* toward this behaviour (A_B) and of the *subjective norms* the behaviour is exposed to (SN_B). Equation (1) represents this influence, where W_1 and W_2 are the personal importance attributed to each component:

The *attitude* is defined by (2) where b_i is the "*belief value*" and e_i is the "*evaluation*" of that belief,

The *subjective norms* is defined by (3) where b'_i is the "*normative belief value*". i.e. the reference belief of the group the receiver considers himself in – and m_i the "*motivation*" to follow the group beliefs.

$$I_B = f(W_1 \times A_B + W_2 \times SN_B) \quad (1)$$

$$A_B = \sum b_i \times e_i \quad (2)$$

$$SN_B = \sum b'_i \times m_i \quad (3)$$

The standard example provided in persuasive communication lecture books ([22] for example) relates to the act of filing and paying taxes. The belief b_i would then be "I should file taxes" and the final intention I_B "I will file my taxes". The usual attitude A_B towards the behaviour is very low as its evaluation in the person's mind is low, however, the social norms, influenced by laws and peer pressure, are high. Thus, at the end, the intention towards the behaviour is still high and the taxes will be filed and paid.

A similar representation of human reasoning was developed within the Belief-Desire-Intention (BDI) model [5]. This model describes the actual intention of realising an action – or a behaviour – that is linked to someone's desires about this behaviour and the relying world representation contained in its beliefs. However, [5] does not provide a model as strict as the one proposed by [1] to describe the relations between these layers of practical reasoning. [6] tries to rationalise this relationship between beliefs and goals – or intentions – in the practical reasoning models close to BDI.

Evaluating a change of behaviour can thus be done, according to this model, by evaluating a change in the beliefs linked to the behaviour or a change in the influences of social norms. In a controlled experiment, one can choose to evaluate one or the other independently.

In particular, in a controlled experiment were the change in social norms' influence is controlled for – on a short term evaluation for example –, researchers can evaluate a change in beliefs and evaluate the persuasion as a change in the attitude towards a behaviour instead of direct behaviourial observation.

Beliefs can be linked to the judgement of a behaviour, but also to some external representation. For example [8] uses such a technique to evaluate the persuasiveness of their embodied conversational agent where instead of measuring the actual buying behaviour to see if the system is persuasive, the authors use a view of the attitude towards this behaviour given by the amount of money participants are ready to spend. However, this measure stays limited to the domain.

In this paper, we discuss beliefs that can be linked to behaviour's intentions as well as to a ranking between a set of items, which we believe can be applied to various domains and can provide a measure for comparison between researches. [18, 3] use the desert scenario task to provide a ranking task to participants: they are told that they are stranded in the desert after a plane crash and should rank a set of items (compass, map, knife, etc.) according to their usefulness for the participants' survival. The resulting ranking provides an external representation of the set of beliefs each participant has formed about the utility of each item.

The ranking does not provide a detailed view on every internal belief that the user holds about the situation, however, if the user changes this ranking, this change represents a measurable change in the internal beliefs. According to the Theory of Reasoned Action this change in beliefs has an impact on the user's intention towards the behaviour, and we can assume that the measured persuasion has an influence on the behaviour too.

3 Measuring Persuasiveness

The ranking task provides an observation of the user's beliefs that can be used to extract a metric evaluation that can be shared and compared between research domains. In this section, we present the general metric measure that can be used and consider different issues in implementing a ranking task and applying the persuasiveness metric.

When participating in a ranking task, the participants first give their preferred initial ranking R_i of items (for example, in the desert scenario task: knife, compass, map, ...) and then engage in with the persuasive system which attempts to change the participants' ranking to a different ranking R_s; at the end of the persuasion session, the participants can change their items choice to a final ranking R_f (see figure 1).

The persuasiveness of the session is measured as the evolution of the distance between the user's rankings (R_i, R_f) and the system's goal ranking (R_s). If the system is persuasive, it changes the user's beliefs about the items ranking towards a ranking similar to the system's ranking. The change of beliefs is reflected by the evolution of the distance between rankings as defined by equation (4).

$$P = \Delta(d(R_f, R_s), d(R_i, R_s)) \quad (4)$$

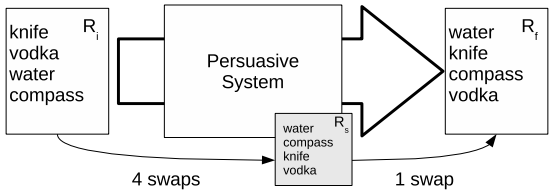

Figure 1. Desert Scenario Ranking Task Example

There exist a number of distance measures between rankings [13, 10, 12]. The Kendall τ coefficient is generally used to measure a difference between rankings. However, this measure is not a metric and is not always straightforward to interpret. A part of the Kendall τ coefficient is however a metric and provides an intuitive measure in the ranking task. The *"Kendall τ permutation metric"* [13] is used to compute the pairwise disagreement between two rankings; measuring the number of swaps between adjacent items to get from one ranking to the other ranking. The Kendall τ permutation metric between the rankings R_1 and R_2 is defined in Equation (5)[2] where P_{airs} is the set of all possible pairs of items of R_1 and R_2.

$$K_\tau(R_1, R_2) = \sum_{\{i,j\} \in P_{airs}} \bar{K}_{i,j}(R_1, R_2) \tag{5}$$

$$\bar{K}_{i,j}(R_1, R_2) = \begin{cases} 0 & \text{if the pair of items } i \text{ and } j \text{ are in the same} \\ & \text{order in the two rankings,} \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

Equation (7) defines the evolution of the Kendall τ permutations metric during the persuasive session and provides a metric evaluation of the system's persuasiveness.

$$\begin{aligned} \bar{P}_{ersuasiveness} &= \Delta(K_\tau(R_i, R_s), K_\tau(R_f, R_s)) \\ &= K_\tau(R_i, R_s) - K_\tau(R_f, R_s) \end{aligned} \tag{7}$$

For example, if the user's initial ranking of the items is $R_i = map > flashlight > compass$ and the system goal ranking is $R_s = compass > flashlight > map$. The Kendall τ permutations metric is calculated with the table of pairs:

R_i	R_s	$\bar{K}(R_i, R_s)$
map > compass	map < compass	1
map > flashlight	map < flashlight	1
flashlight > compass	flashlight < compass	1
	$K_\tau(R_i, R_s)$	3

If the final user ranking is $R_f = flashlight > compass > map$, the table of pairs is:

[2] from [10]

R_f	R_s	$\bar{K}(R_f, R_s)$
compass > map	compass > map	0
flashlight > map	flashlight > map	0
flashlight > compass	**flashlight < compass**	**1**
	$K_\tau(R_f, R_s)$	1

At the beginning of the persuasive session, the distance is maximum between the two rankings – three swaps are needed – whereas, at the end of the session, only one swap is required. The persuasiveness metric is then: $\bar{P}_{ersuasiveness} = 3 - 1 = 2$.

For an n items ranking, the range of the persuasiveness metric is thus

$$[-\frac{n \times (n-1)}{2}, \frac{n \times (n-1)}{2}]$$

To be able to compare different persuasive systems that can rely on heterogeneous ranking task with different numbers of items, we need to normalise this persuasiveness measure as defined by equation (8).

$$P_{ersuasiveness} = \frac{2 \times (K_\tau(R_i, R_s) - K_\tau(R_f, R_s))}{n \times (n-1)} \tag{8}$$

3.2 Interpretation and Constrains

In this general approach to the ranking task, the normalised persuasiveness metric will have a minimum of -1 and a maximum of +1.

- The *minimum* corresponds to the case where the participants actually made the maximum number of swaps **away** from the system's ranking between the initial and the final ranking.
- A *null* $P_{ersuasiveness}$ means that the participant did not change the ranking and that the system was not persuasive.
- The *maximum* $P_{ersuasiveness}$ corresponds to a successful persuasion of the system as the participants will have done the maximum number of swaps **towards** the system's ranking and $R_f = R_s$.

In this general setup of the ranking task, there is however a issue for the interpretation of the results. What does it mean for the persuasive system that the users change their beliefs **away** from the persuasive goals that the system was seeking? Was the system extremely bad? is $P_{ersuasiveness} < 0$ worst than $P_{ersuasiveness} = 0$? It is actually difficult to interpret the $P_{ersuasiveness}$ metric in its negative range.

[9] discusses *"arguments that backfire"*, where the use of fallacy lowers the audience's trust in the speaker and thus lowers the effectiveness of the argumentation. This might make the whole persuasion "backfire", yielding negative results that will make the audience go against the speaker persuasive goals, even if they shared initial beliefs. This will explain negative $P_{ersuasiveness}$ results as the shared beliefs represented in the initial ranking will be lost and the participant will provide a final ranking further away from the system's goal ranking than the initial ranking. The negative results are thus valid in their interpretation and can help detect backfiring argumentation strategies that alienate the audience.

However, an additional issue with this setup of the ranking task makes it hard to compare between different domains instantiation. For example, in an extreme case of this general view of the ranking task, the user can enter an initial ranking R_i that is the same as the ranking R_s chosen by the system. In this case, the task of the persuasive system is not to persuade users but to keep the same ranking and

the only evolution of ranking that can be observed are swaps **away** from the system's ranking. In this extreme case, it is not interesting to compare the persuasiveness metric with another persuasive session where $R_i \neq R_s$ and where the system would have had to actually do some persuasion *effort*.

To be able to compare different persuasive systems with this ranking task, the persuasion task, with regards to this ranking task, should be of comparable *effort*. Normalising the $P_{ersuasiveness}$ allows to compare different persuasive task that have a different number of items, but does not protect from comparing a system persuading the user to do a little relative number of swaps with a system that has to persuade the user of a large relative number of swaps.

A solution to get a uniform $P_{ersuasiveness}$ metric, which can be compared between systems, is to guarantee that each system will have a comparable persuasive *effort*. This can be guaranteed by choosing the system's goal ranking R_s to always maximise the persuasive *effort* by maximising the number of swaps needed to go from R_i to R_s. This is guaranteed by choosing R_s as the invert ranking of R_i as shown in the example given above. In this case, the initial distance between rankings is $\frac{n\times(n-1)}{2}$ where n is the number of items in the ranking.

If the ranking task is defined with this constrain, then we can write the $\bar{P}_{ersuasiveness}$ as defined by equation (9) which implies that the persuasiveness range is $[0, \frac{n\times(n-1)}{2}]$ and the normalised persuasiveness, defined by equation (10), has a range of $[0, 1]$. If the participant is not persuaded by the system, then $R_i = R_f$ and $P_{ersuasiveness} = 0$ but if the system is persuasive, then the participant has done the maximum number of swaps **towards** the system ranking and $P_{ersuasiveness} = 1$ as $R_f = R_s$.

$$\bar{P}_{ersuasiveness} = \frac{n \times (n-1)}{2} - K_\tau(R_f, R_s) \quad (9)$$

$$P_{ersuasiveness} = 1 - \frac{2 \times K_\tau(R_f, R_s)}{n \times (n-1)} \quad (10)$$

When designing the persuasive experiment and setting the ranking task, the researcher should therefore be very attentive that the chosen system's goal ranking is always the invert of the user's ranking. The system must also be able to achieve such a persuasion.

The non maximised setup of the ranking task is helpful in detecting "backfiring" argumentation which will move the user's beliefs away from the system's goal belief. This provides a good insight of the argumentation process but is not usable for comparing different systems' performances to change the user's belief. The second measure, can be used for this purpose as it guarantees the maximisation of the persuasion, however, nuances of the belief change will be lost as, in this setup, there is no option for the participant to *disagree* more with the system. The goal of the experiment should thus set the measure to use:

- if the experiment is designed to evaluate the persuasive strategies of the system, then it is interesting to leave space for the participants to *disagree* with the system and the first measure should be preferred.
- if the experiment is designed to compare the system's effectiveness to change the user's beliefs between system, then it is recommended to use the second "maximised disagreement" measure that removes the bias of the initial belief choice.

4 Sample Experiment and Results

[18, 3] used the desert survival scenario [14] ranking task to evaluate the persuasiveness of dialogue sessions but did not formalise a general persuasiveness metric. In our research, we have used a similar ranking task based on a different scenario to evaluate a persuasive system with the formal $P_{ersuasiveness}$ metric described earlier. In this section, we report initial observations on the use of this metric as well as an example of a different scenario where the ranking task can be used.

Our research was evaluating a human-computer dialogue system able to discuss with users to persuade them. The domain chosen to evaluate this dialogue system was similar to a restaurant recommendation scenario. Twenty-eight participants were told that they would discuss with an automated dialogue system simulating one of their friend in a chat about a shortlist of restaurants where they could bring mutual friends for dinner.

After having been explained the scenario, the participants are presented with a list of ten restaurants described by five attributes (food quality, cuisine type, cost, service quality and decor) and are asked to choose three restaurants they would like to propose as possible alternatives to their group of friends. They can choose any three restaurants and rank them in their order of preference.

The actual dialogue system has access to a database of around one thousand restaurants[3], but asking the user to evaluate, in a short time, all of these restaurants is not realistic. In the same way, asking them to rank the full list of ten restaurants is not possible and would not correspond to a natural task that the participants would perform in real life.

After having given information about their restaurants preference and a specific restaurants choice, the participants are faced with a dialogue session with a system simulating a friend that tries to persuade them to keep the same selection of three restaurants, but to choose a different preference order. In this case, to ensure maximum persuasion *effort*, the system always chooses a ranking of restaurants that is the invert of the user's choice.

At the end of the dialogue, the participants are asked to give their final ranking of restaurants reflecting their agreement with the simulated friend. This is used as the final ranking to measure the persuasiveness of the system. The participants are also asked to fill in a questionnaire relating to different aspects of the dialogue system. In this experiment, to evaluate the fitness of our evaluation metric, the participants are asked to rate the statement "*The other user was persuasive*" on a five points likert scale: "*Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly Agree*".

This statement evaluates the persuasion perceived by the participants during the dialogues. The persuasiveness metric applied in this case shows that there is a significant correlation between the user perception and the persuasion measured through the ranking task (Spearman $\rho = 0.70$, $p < 0.01$)[4]. This confirms that the measure is at least as good as asking the question directly to the users. However, getting such direct measure might bias the answer of the users.

Observation of the answers from the user also shows the need for a side measure of persuasiveness. In the seven participants that answered that they "neither agree nor disagree" to the statement, an outlying participant that does not perceive a strong persuasion but is still persuaded more than the other. In this case, the side measure of

[3] provided by M.A. Walker from [23]

[4] in a similar setup with 52 participants, the same question was asked and also yields a significant correlation with the persuasiveness measure (Spearman $\rho = 0.38$, $p < 0.01$)

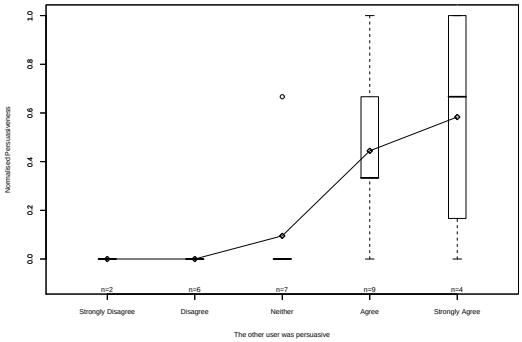

Figure 2. Correlation Between the Perceived Persuasion and the Measured Persuasion. n is the number of participants that gave this particular answer.

the rank change allows to see that users were persuaded even if they did not perceive a strong persuasion from the system.

Similarly, the "strongly agree" answers show that there is a distribution of the persuasiveness measure along the whole axis: some of the participants that perceived a strong persuasion from the system did not actually change their ranking[5]. This illustrates the case where users do actually feel that they are persuaded but might not have changed their beliefs accordingly. In which case, the system cannot be said to be persuasive.

Thus, a side measure of persuasion, that does not directly rely on participants' self evaluation can show more information about the actual persuasion process while staying a good indicator of the system's persuasive performances.

5 Ordering of Beliefs vs. Ordering from Beliefs

In belief revision literature, in particular within the AGM model [2], someone's belief set is represented as a set of consistent axioms on which operations can be performed to revise or update the beliefs. Axioms can be added or removed from the set at each revision to maintain the consistency of the belief set. However, in someone's mind, not all beliefs are equal as some are said to be more *entrenched*, and they are harder to remove from the person's belief set.

This entrenchment affects the possible revisions of the belief set and can be seen as a preference ordering of beliefs in the person's mind. Beliefs higher in one's preferences will be harder to change and remove from the belief set than lower beliefs.

Belief revision, which is the base of the persuasion discussed in this paper is thus seen as an operation on a set of ordered beliefs that can be extended, reduced or reordered. This *ordering of beliefs* could be seen as similar to the ranking task proposed in this paper: the ranking represents the entrenchment ordering of the user's belief and the system's task is to make the user revise such ordering.

However, the ranking task is actually less abstract and each item of the ranking does not need to directly map to a belief in the participant's mind. For example, in the ranking task of the desert survival scenario, each item does not map to one of the participant's belief (or an axiom representing such belief).

For instance, two items are available in the desert survival scenario: an airmap and a compass. Most participants have the belief that they can use these two items to find their way out of the desert. If these two items are ranked high in the initial ranking, the system can assume that the participant holds the following beliefs:

- "I can walk across the desert to find rescue."
- "I can find my way to rescue on the map."
- "I can use the compass for orientation on the map."

The ranking of the compass and the map over a flashlight for example does not represent a direct preference ranking over beliefs but that the participant sees more use for these items than for the flashlight, *because* of his current beliefs.

In the restaurant domain, the ranking represents the users preference towards the restaurants, these preferences are not a direct mapping to an entrenchment ordering, but is still related to this concept. If a user ranked a *Pizzeria* over a *Grill*, this might map to a set of preference ordering over the cuisine type. However, it might also be that the Pizzeria is cheaper than the Grill.

Another example is the smoking cessation program, the ranking items could be directly mapped to a set of beliefs related to why the participant is smoking, such as: "smoking makes me feel better", "smoking makes me look cool", "smoking will kill me", etc. However, these might be hard to change as some of the beliefs might be too entrenched. A different, indirect, ranking task could evaluate the change of beliefs about smoking while avoiding too much entrenchment bias; for example, the participants could be asked to rank a set of items they would buy first if they had a limited amount of money, such a set could contain "a bottle of water", "a pack of cigarettes", "a newspaper", "a lighter", etc. The reranking of the items relating to smoking, while not ensuring that the participants will stop smoking, will still show a change in their attitude towards the smoking behaviour.

The choice of the ranking items should thus not be directly mapped to a set of ordered beliefs or preferences, but to a set of items that represent, in practice, a set of knowledge and of preferences about the domain. The ranking will be guided by the user's belief: a *ranking from beliefs*, but might not directly map to the *ranking of beliefs* in the user's mind.

6 Conclusion and Discussion

In this paper, we have introduced the different approaches of evaluating systems' persuasion through the state of the art of automated persuasion. We have also formalised a framework providing a reusable persuasiveness metric that could be used by other researchers to compare different automated persuasion approaches.

The applications illustrated in the paper are short term setups that cannot evaluate the long term impact on the participants, but actually, this ranking task can also be used as a measure in long term evaluations. For example, in the case of the smoking cessation problem [21], the use of a ranking task might have provided more insight in the beliefs change of the users after the first intervention; six months later, the same ranking task without extra intervention might have been used to evaluate the beliefs that remained of the persuasion, even if the participants did not stop smoking. Such ranking task would have thus given more insight on why the system was not effective.

This paper provides sample results that show that the proposed persuasive measure is at least as good as directly asking the user about the persuasion while providing a *hidden* measure that does not bias the participants. It remains to be shown if this measure correlates with actual behaviour change. This was not in the scope of our

[5] note that this could also be due to a misunderstanding of the instructions by some of these participants.

research but will have to be evaluated if researchers have to use the measure in more complex domains where the user's personal beliefs might not have a strong weight in comparison to the social norms affecting the behaviour.

In addition, the change in the ranking evaluates a change in the user's attitude but might not be directly linked to persuasion as such a change might come from coercion and threats. Thus a measure of coercion is required to ensure that the change measured by the proposed metric comes actually from persuasion. For example, to evaluate coercion, in the reported sample experiment, the user was directly asked the question: "The other user was not forceful in changing your opinion" which did not show to correlate with the persuasiveness metric.

We have shown that the ranking task can be applied to different domains, however, to use such a task, the persuasion must be performed on a domain where behaviours or attitudes can be mapped to a ranked set of items. It is clear that not all persuasive domains can be reduced to a ranking task. In addition, doing such reduction might limit artificially the scope of research on automated persuasion.

We believe that the formalisation of the ranking task as a framework for evaluating persuasive systems is a first step towards finding an appropriate evaluation methodology for comparing persuasive systems. It is important for the development of the field of automatic persuasion and natural argumentation that researchers extend their work on a set of standard evaluation frameworks that can be used to evaluate and compare systems on long and short term changes in the user's beliefs, attitude and behaviours. In addition, this paper only discussed the problem of evaluating the existence and ranking of beliefs linked to a behaviour, but the problem remains to find a task to evaluate the social norms influencing the behaviour.

REFERENCES

[1] I. Ajzen and M. Fishbein, *Understanding attitudes and predicting social behaviour*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.

[2] C. E. Alchourròn, P. Gärdenfors, and D. Makinson, 'On the logic of theory change: Partial meet contraction and revision functions', *Journal of Symbolic Logic*, **50**, 510–530, (1985).

[3] Pierre Andrews, Suresh Manandhar, and Marco De Boni, 'Argumentative human computer dialogue for automated persuasion', in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pp. 138–147, Columbus, Ohio, (June 2008). Association for Computational Linguistics.

[4] Timothy W. Bickmore and Rosalind W. Picard, 'Establishing and maintaining long-term human-computer relationships', *ACM Transaction of Human-Computer Interaction*, **12**(2), 293–327, (June 2005).

[5] Michael E. Bratman, *Intention, Plans, and Practical Reason*, Cambridge University Press, March 1999.

[6] F. Paglieri C. Castelfranchi, 'The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions', *Synthese*, **155**, 237–263, (2007).

[7] Giuseppe Carenini and Johanna D. Moore, 'An empirical study of the influence of argument conciseness on argument effectiveness', in *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, ed., Hitoshi Iida, pp. 150–157, Hong Kong, (October 2000).

[8] J. Cassell and T. W. Bickmore, 'Negotiated collusion: Modeling social language and its relationship effects in intelligent agents', *User Modeling and Adaptive Interfaces*, **13**(1-2), 89–132, (February 2002).

[9] Daniel H. Cohen, 'Arguments that backfire', in *The Uses of Argument: Proceedings of a conference at McMaster University*, ed., David Hitchcock, pp. 58–65. Ontario Society for the Study of Argumentation, (April 2005).

[10] Ronald Fagin, Ravi Kumar, and D. Sivakumar, 'Comparing top k lists', in *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 28–36, Philadelphia, PA, USA, (2003). Society for Industrial and Applied Mathematics.

[11] M. P. Jensen, W. R. Nielson, J. M. Romano, M. L. Hill, and J. A. Turner, 'Further evaluation of the pain stages of change questionnaire: is the transtheoretical model of change useful for patients with chronic pain?', *Pain*, 259–264, (June 2000).

[12] M. G. Kendall, 'A new measure of rank correlation', *Biometrika*, **30**(1/2), 81–93, (June 1938).

[13] Maurice Kendall and Jean D. Gibbons, *Rank Correlation Methods*, A Charles Griffin Title, fifth edn., September 1990.

[14] J. C. Lafferty and P. M. Eady, *The desert survival problem*, Plymouth, Michigan: Experimental Learning Methods, 1974.

[15] K. Lorig, R. L. Chastain, E. Ung, S. Shoor, and H. R. Holman, 'Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis', *Arthritis and rheumatism*, **32**(1), 37–44, (1989).

[16] *Symposium on Persuasive Technology, in conjunction with the AISB 2008: Convention Communication, Interaction and Social Intelligence*, eds., Judith Masthoff, Chris Reed, and Floriana Grasso, Aberdeen, April 2008.

[17] Gr Miller, 'On being persuaded: Some basic distinctions.', in *Persuasion: New directions in theory and research*, eds., M. E. Roloff and G. R. Miller, 11–28, SAGE Publications, (January 1980).

[18] Youngme Moon, 'The effects of distance in local versus remote human-computer interaction', in *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 103–108, New York, NY, USA, (1998). ACM Press/Addison-Wesley Publishing Co.

[19] *Persuasive Technology. Proceedings of the Third International Conference, PERSUASIVE 2008*, eds., H. Oinas-Kukkonen, P. Hasle, M. Harjumaa, K. Segerståhl, and P. Øhrstrøm, volume 5033 of *Lecture Notes in Computer Science: Information Systems and Applications, incl. Internet/Web, and HCI*, Springer, Oulu, Finland, July 2008.

[20] J. O. Prochaska and Carlo Diclemente, 'Stages of change in the modification of problem behavior', *Progress in Behavior Modification*, **28**, 183–218, (1992).

[21] Ehud Reiter, Roma Robertson, and Liesl M. Osman, 'Lessons from a failure: generating tailored smoking cessation letters', *Artificial Intelligence*, **144**(1-2), 41–58, (2003).

[22] James B. Stiff and Paul A. Mongeau, *Persuasive Communication*, The Guilford Press, second edn., October 2002.

[23] M. A. Walker, S. J. Whittaker, A. Stent, P. Maloor, J. Moore, M. Johnston, and G. Vasireddy, 'Generation and evaluation of user tailored responses in multimodal dialogue', *Cognitive Science: A Multidisciplinary Journal*, **28**(5), 811–840, (2004).

Persuasion at the Museum Café: Initial Evaluation of a Tabletop Display Influencing Group Conversation

Cesare Rocchi, Oliviero Stock, Massimo Zancanaro, Fabio Pianesi and Daniel Tomasini [1]

Abstract. A café table is a traditional setting for conversation. Tabletop displays may have an active role in this connection. In particular for a museum scenario, conversation after the visit is important for a joint elaboration of a small group visit experience. We propose the museum café as the location to introduce a tabletop display meant to foster and support conversation about the visit. The goal of the system is to influence the development of the conversation by adopting persuasion techniques. We describe a system that monitors the conversation among the visitors and dynamically shows visual stimuli on the table surface. An initial formative evaluation is conducted through a series of qualitative user studies.

1 INTRODUCTION

In museum scenarios, informal conversations among small group of visitors play a fundamental role in the learning process, as ethnographic studies have clearly demonstrated [4]. We take this as the inspiration for this work. How can technology induce people to entertain a conversation about their experience at the museum and help sustain it? This question is closely related to the work of Fogg [2], which highlighted the potential of computers as persuasive tools that can influence people's behaviour, also in an educational entertainment scenario like the one we propose.

Most of the current technology for museum visits addresses the single user [3]; people, however, tend to visit a cultural site with families, groups of friends, etc. Petrelli and Not report that 45% of the visitors go in organized groups [4]. Mobile guides and kiosks thus are in risks of hampering rather than fostering conversation. We propose a novel aspect: technological tools that provide support after the visit, when visitors can have a conversation about their experience. In particular we investigate a tabletop application placed in the museum café specifically designed to influence the subject of the conversation and the behaviour of the group.

The table is instrumented with sensors and its top surface is used as medium to display persuasive messages aimed at influencing the conversation of the group. Conversation is tracked through word spotting [5], and the group's behaviour is monitored. Reasoning about the overall conversation configuration and the visit to the museum permits to drive the system actions: the system chooses specific presentation strategies that lead to specific output on the tabletop.

The scenario, at the museum café after the visit, includes three phases:

a) a phase where the system promotes a conversation about the museum visit experience with the goal of shifting the group discussion into a specific topic of the cultural experience;

b) a phase that supports conversation by providing content appropriate to the specific topic being discussed and the state of the conversation;

c) a phase where one member or the whole group explicitly seek further information about some cultural heritage topics by interacting with the system.

In the present paper we concentrate on the first two phases, where there is no explicit input to the system by the participants; the system observation of the non mediated participants' interaction is used as a sort of implicit input. When people entertain a conversation on a topic not related to the visit, the system tries to influence the conversation by attracting them towards visit-related topics, with techniques reminiscent of the tradition of advertising. When the conversation is about a museum topic, it supports it by proposing relevant material, also drawing on information about their visit history.

In the following, after a short review of related works, we report about an initial Wizard of Oz user study investigating subjects' reactions to the tabletop display during a conversation after a museum visit. The study explores the effect of a number of communicative strategies exploited by the system, which are borrowed from semiotics and advertisement techniques. The analysis of video recorded data and post-study interviews has helped defining the technological requirements. The actual architecture implementing those requirements is then described in the following section. Presentation strategies are specifically in focus in the following part, with the accent put on the persuasive connotation. We then describe the present implementation and briefly discuss it.

2 RELATED WORKS

The system proposed in this paper has some affinity with a peripheral display [6] in that the table is not central to the attention of the group and people may look at it only occasionally. Yet peripheral displays are normally used as secondary sources of information, separate from a user's primary, focal task [7] and are usually meant to have a passive role and just aim at making users aware of easily graspable

[1] FBK-irst, 38100 Trento, Italy.
Email: {rocchi/stock/zancana/pianesi/tomasini }@fbk.eu

information such as weather or stock graphics. On the opposite, our system actively monitors the group's behavior with the intention to induce specific behavior by displaying information when contextually appropriate.

There have been many studies on the display of information through social tools. For example, the Tangible Bit project was about conveying information to increase awareness of people's presence and activity [8]. Another example is Groupcast, a wall projected office application that creates informal interaction opportunities by displaying mutual interest to people passing by [9]. Drift is an interactive table that displays an aerial photo of England through a hole, to foster interpretation and engagement [10]. Qualitative observations showed that people got engaged by interacting with the system and narrating about the places spotted. Hello Wall is a digital wall made of a grid of lights [11]: depending on people distance, the wall changes communicative function (ambient, notification, interaction). Abstract light patterns convey information about mood, presence and crowdedness.

DiMicco and Bender [12] have experimented with a system that, by monitoring working groups, presents information about relational behavior in the form of graphical displays on a tabletop device, to affect group behavior.

A similar approach was pursued by Sturm and colleagues [13] who used a tabletop device as a peripheral display aiming at the same self-regulatory effect as discussed above. In their approach, they display not only the speaking time but also the gaze behavior of their participants. Their results show a similar effect of Dimicco and colleagues for what concern the speaking behavior and no effect on gaze behavior.

Kim and colleagues [14] used a portable device called the sociometric badge to monitor speaking activity and other social signals in a team. They report a graphical representation of the group behavior on a private display. Their results showed a reduction of the overlapping speech but not a significant increase in solo speech.

All these approaches are based on the idea that reflection on one's own behavior may bring to rational decisions about behavior changing [15]. Usually these systems are applied in a team-work scenario where each participant is motivated to achieve his/her goal, e.g. a successful meeting and/or a well accepted personal appearance. Their approach, focused on balancing the contributions of the participants, has been proven to be effective in reducing the involvement of dominant participants but not in increasing the participation of the less active ones.

We propose a different approach: our system intends to affect the group behavior by presenting on a shared interface (namely the café table) contextually appropriate visual material in a novel way, reminiscent of the tradition of advertisements: attention catching, evocative and cognitively stimulating.

Our approach is motivated also by studies in the field of persuasive technology [2]. Fogg identified seven strategies for persuasive technology tools:

- Reduction: making something complex appear simpler.
- Tunneling: demand to an expert.
- Tailoring: providing relevant information.
- Suggestion: act at the right time with a message.
- Self-monitoring: tracking the desired behavior
- Surveillance: publicly observe one's behavior
- Conditioning: reinforce target behavior with positive "reward".

With reference to the above strategies, we propose a system that features suggestion and conditioning. Suggestion is based on interventions at the right time to maximize the effectiveness of the persuasive message. A suggestion-based technology actively induces someone to do something she might not have done otherwise. In our case, the system shows stimuli meant to support the current activity of the group (a conversation about the visit experience) or favor a behavioral change (make some participant more active). Conditioning is based on the provision of positive feedback to favor the persistence of an already occurring behavior. This strategy is usually adopted when the system aims at supporting an ongoing conversation. Our system also uses a tailoring strategy, which appropriately selects content according to the topic currently discussed. The stimuli presented by this strategy are related to the current topic discussed, in a way similar to recommender systems [16].

Data about the conversation are processed by the system to output stimuli that realize suggestion, conditioning and tailoring strategies In a nutshell, instead of revealing to the group the social dynamics and requiring them to take into account the information and act rationally to achieve a given meeting goal, we aim at directly influencing (modifying or sustaining) the behaviour of the group.

3 THE INITIAL STUDY: WIZARD OF OZ

A Wizard of Oz experiment was initially performed to study the reaction of the users to an active table in the museum café (see [17] for the details). We hypothesized that data available to the system are: images and texts about the exhibition, profiles and visit's logs for each visitor and an automatic speech system able to understand the topic of the conversation.

In this study three groups of 4 people were invited in our lab to visit a reconstruction of the "Cycle of the Months" frescoes in Torre Aquila, (Castello del Buonconsiglio), Trento, Italy. Subjects were given a four-page booklet to help them during the visit and were told that the purpose of the study was to test the content of the booklet. After the visit people were conducted to another room and were invited to sit at a table while waiting for the experimenter to come back. The wizard, located in another room, monitored the group behaviour and controlled the presentation of visual stimuli projected onto the table. After the study, an experimenter debriefed the group about the real purpose of the study and conducted a semi-structured interviewed aimed at eliciting subjective impressions.

Recorded sessions and interviews have been analysed addressing the following questions:

- Did a stimulus catch the attention of one or more users?
- Did a particular system action (e.g. zooming on an image) favour the change of a topic?
- Did a graphical effect upset the users?

The questions of the interview addressed the role of images and words, the density of stimuli displayed and the conceptualization of the system's behaviour in general.

From the observations and the interviews, it resulted that the system was recognized as a useful tool to wrap up a visit, especially in case people were not acquainted with the exhibition. Subjects also reported the feeling that the table sometimes 'follows' the conversation and tries to propose new hints. They also said to be upset in case of weird behaviour, especially when the image supporting the conversation disappears. All the groups reported that when the discussion of a topic was languishing they used the stimuli on the table to start a new conversation. Yet graphic-intensive effects like pulsing and flashing have been considered too upsetting, especially when there is an ongoing conversation.

Figure 1. A snapshot of the Augmented Café Table

4 THE SYSTEM

Starting from the insight from the WoZ we developed the Augmented Café Table, a system that analyses the conversation of a group of people around the table and presents a set of stimuli in order to change or sustain the conversation. The Augmented Café Table is a tabletop display with the form factor of a café table (see Figure 1). At present, the interface is top-projected, for future releases we will experiment with back-projection and multi-touch capabilities. In the final scenario, people are sitting at a museum's café table after having visited the frescoes. The current system targets the "Cycle of the Months" frescoes referred to above. The "desired" topics for conversation are the frescoes themselves, and the set of stimuli exploited by the system includes images and videos of the frescoes or of related details thereof, as well as short sentences relevant in the domain.

The system employs a set of microphones to capture the users' conversation, which is analysed using a keyword spotter. Knowledge about the behavior of each individual in the museum is also traced using the visit's logs from a multimedia guides. Logs provide information about the exhibits visited, and the amount of information the guide provided. As said above, the system has two roles:

a) the system promotes conversation about the museum visit experience trying to shift the discussion towards a specific visit-related topic;

b) the system provides contents appropriate to the current topic of discussion and the state of the conversation.

The interface displays visual stimuli such as floating words and pictures meant to be cues for the conversation, whereby the group can discuss ideas, share impressions, exchange opinions and, in general, get along with the spirit of the visit.

The system is organized along three modules: perception, interpretation and presentation (see Figure 3).

The perception module receives and processes data from sensors. The first type of data relates to voice activity, captured by microphones. The perception module includes a Voice Activity Detector and a keyword spotter, which recognizes words uttered during the conversation. The output of the keyword spotter is a series of words, with attached a measure of confidence indicating the reliability of the result. A second type of data comes from the visual scene and where face detection mechanism deals with visual attention toward the table. The input from four webcams is processed by a Haar-based head detector [18]. The visual attention of each participant is estimated by calculating changes in the proportion of the bounding box of the face: when the box is vertically squeezed, the system detects a visual attention toward the table.

The interpretation module analyses the state of the conversation, which is modelled along two dimensions: state of the group and state of the table. The first dimension relates to the content actually discussed and the behaviour of the participants around the table, while the second deals with the actions of the table itself. One of the features of the conversation state is the topic currently discussed. Topic detection is based on a semantic network as in Figure 2. Each node of the network is a topic, which represent relevant concepts related to the visit. Our network includes one topic for each panel of the fresco, plus general topics which are shared by more than one panel, e.g. life in the middle age or information about the restoration. Each topic has attached a set of keywords, which describe the node.

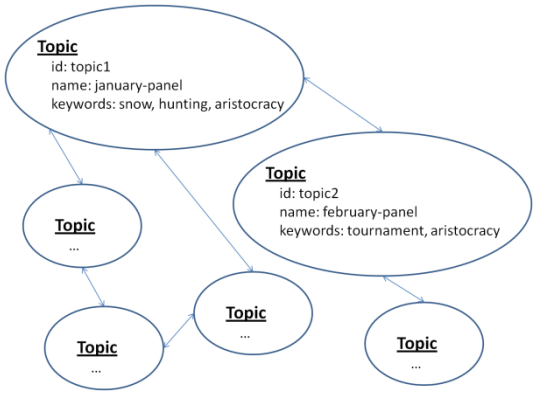

Figure 2. An excerpt of the topic network.

Such model allows the system to compute topics' "connectedness". This information is then processed by the strategy selector to dynamically propose new topics to be discussed. In doing so, the system takes into account the history of the conversation. The elements of the history are topics, with attached information about the duration (how long a topic has been discussed) and the list of speakers who contributed to the

discussion. In case the system has not recognised anything pertaining to one of the expected topics, it just records information about participants' activity and the topic of the conversation is marked as "out of domain" assuming that the group is conversing on something unrelated to their museum experience. The history of the conversation is used by the interpretation module by considering also its evolution in a given time window:

- level of participation . E.g., a person has been too much or too little active.
- conversation development. E.g., the recent conversation has been "jumpy", i.e. recently discussed topics are unrelated one to the other.
- topic coverage. E.g., the current topic has already been discussed extensively.

The information discussed above is cumulatively used by the system to trigger a reasoning mechanism which selects an appropriate presentation strategy to be realized on the tabletop surface (see below). Beside the current state, the choice of a presentation strategy is also conditioned by the presentation currently displayed on the surface and the history of strategies previously used.

For example, to support continuation of the conversation on the same or related topics, and enforce the cohesion of the tabletop dynamics, the system will reason on the topic network and both on the history of the conversation and the history of its own presentations.

Animation of images and words is a key characteristics of the stimuli displayed on the table. Motion captures attention and is easier to identify in the periphery than color and shape [19]. A proper timing of animations is indeed of paramount importance since the onset of motion is more effective at capturing attention than motion itself [20]. We think that some of the design dimensions commonly adopted by peripheral displays are useful to structure the presentation layer. We consider three dimensions: data representation, notification and transition.

Data representation refers to the way stimuli are shown and the potential impact they can have on the conversation. Since we want to immediately influence the development of the conversation we use images and words as possible stimuli to foster and support a conversation about the visit. The goal of this choice is twofold: to allow focusing on particular aspects of the painting and to foster the visitors' interpretative engagement. By interpretive engagement we mean the attitude of asking questions like: "What is this? What is my experience about it? What can I share with others about it?"

Notification relates to the dynamics the system adopts to show visual stimuli. The way stimuli appear and move is meant to catch the attention and potentially change the behavior of the group. A stimulus appearing or getting larger can have diverse effects on people's perception: it can be simply change blind (a person viewing the visual scene does not detect large changes in the scene) attention grabbing, it can increase awareness, or even directly demand physical action by the user. The notification layer can be thought as a sort of rhetoric of information presentation. The display actions we implemented are meant to obtain different effects in dependence of the state of the conversation. For example, a way to change the topic is to grab the attention of a group member on a detail. In this case the notification has to be clear and indicate a passage: for example only one stimulus is visible and the image is progressively enlarged. On the other hand, if the goal is to introduce a topic related to the current one, the change can be smooth—e.g., a new detail is displayed via a slow fade among already present stimuli. Notification is related also to the visual patterns which objects can be organized into. For example a strategy to notify that certain objects are related exploits the metaphor of spatial proximity, that is more related objects are located closer. In our system we implemented a notification pattern that aligns stimuli along a circle and makes them orbit around a common centre, thus forming a cluster that moves as a block.

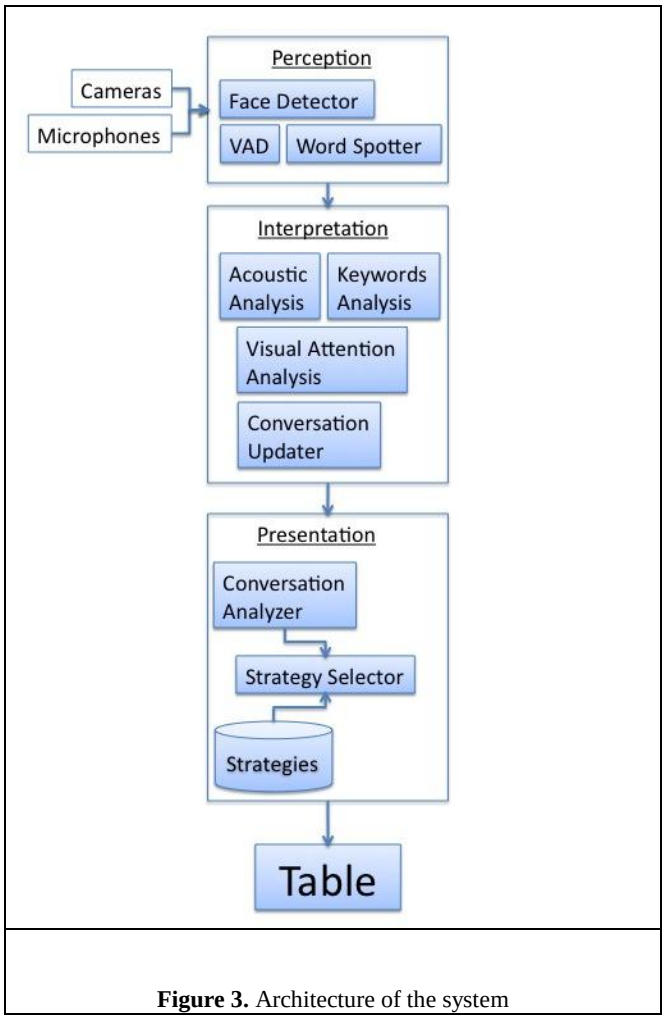

Figure 3. Architecture of the system

The *transition* dimension is related to the notification one. Every action that changes the current visual state of the display can be considered a transition. Transitions exploit graphical effects to attract an appropriate amount of attention from the users and affect the development of the conversation. For example, a topic shift can be suggested by having objects related to the old topic disappear and objects related to the new one

progressively appear; to represent that a topic *t* is proposed as the prevailing one, the object related to *t* can be progressively enlarged while the others gradually scale down.

5 PRESENTATION STRATEGIES

We devised a set of presentation strategies to dynamically select and move stimuli. Strategies are organized according to the three dimensions presented above. The selection of strategies is based on the current perceived state. Each strategy has a goal. Goals implemented at the moment are: (i) support the current conversation, (ii) engage a member in a conversation, (iii) start a conversation.

The first strategy is an implementation of Fogg's conditioning strategy. It happens when somebody is speaking and most people are looking at the table. Here we assume that there is an ongoing conversation and the system supports it. Data represented can be either images or words, animations are slow because the notification level is low, transitions are smooth, to not demand too much attention.

The second type of strategy has the goal to engage a "passive" member in a conversation. This strategy is applied when only one's attention is directed to the table, regardless there is an ongoing conversation. The core of the strategy is to show a stimulus directly toward the "passive" member, in order to suggest a possible topic to discuss. This strategy selects only one image to be shown, for it is more suggestive than a word. The notification level is medium because her attention is already directed toward the table and the goal is to make her aware of a possible topic to be discussed. Transition is medium, because the user has to notice the difference with respect to the previous state of the tabletop.

The third strategy, as the second, is an implementation of Fogg's suggestion strategy. It is applied when the conversation stalls and the goal is to engage the group to discuss a topic related to the visit. It is realized by showing an enlarged image, possibly with many details, which represent possible topics to discuss. The notification of this strategy is very high, for people's attention has to be directed toward the table. The transition level is also high, to ensure that at least somebody notice the change.

Finally, a different type of strategy is based on the idea of using humor to trigger interest. It consists of displaying a fun verbal expression followed, after a short time by an image that refers to it in some way. Like in many broadcast ads we see today on newspapers, wall or TV, slight variations on well known linguistic expressions surprise the audience and get its attention. Normally it is a form of irony that plays on the substitution of an element in the expression with a word that evocates the concept that the ad intends to promote. An example for our case is “Saturday knight fever”, a variation on the well known movie title that evocates the scene of the festive tournament, part of one of the paintings in the museum we are conducting the experiments in. For this study, the humorous expressions were compiled by hand, but see [10] for a system that automatically produces such puns taking into account the context.

6 USER STUDIES

We conducted an observational study on a prototype which implements a subset of the features presented above. As said, the current system works on the scenario of a visit at the “Cycle of the Months” frescoes. This artwork consists of eleven side-by-side frescos each one measuring on average 2 meters wide and 3 meters high, and representing a particular calendar month. The frescos were painted during the 15th century and illustrate the activities of aristocrats and peasants throughout a full year. The main topics for conversation are therefore the eleven frescos (named after the months they represent) and the keywords that can be recognized by the word spotter are 50 words for details and objects depicted in the frescos. The image repository includes images of the full frescos and of relevant details.

For practical purposes, the studies were run at our labs, using two rooms: in the first the Torre Aquila exhibition was partially reproduced; the second room was used as the post-visit meeting place and it was equipped with a table together with top projection, and a camera to record the sessions.

The subjects in a group of 3 or 4 were initially welcomed and asked to visit the reconstruction of the frescoed room. Each of them received a booklet describing the frescoes. To induce a controlled difference in the experiences, two subjects received a booklet slightly different from the others: one contained more details about some of the frescos and the other some information about the restoration process. In order to reduce the bias and to avoid getting to much attention to the table since the beginning of the experiment, the subjects were told that the purpose of the study was to assess the quality of the information provided in the booklet. After the visit, the subjects were accompanied in the other room, where the only available piece of furniture was the system table, and asked to wait for the experimenter to come back. Soon after, the table started to display stimuli according to the phases described above. After approximately ten minutes the experimenter came back and the subjects were debriefed with a short unstructured interview about their experience with the table; the topic of the interview were people’s feelings and attitudes towards the table, the way it functioned, and its place in a real museum.

A total of 5 groups participated in the study; subjects were balanced with respect to gender. Their ages range from 35 to 45. They were all volunteers and, with the exception of two computer programmers, the others had no specific technical skills. Few of them had been in Torre Aquila. All the sessions were video recorded.

The following discussion is based on qualitative observations of the videotapes of the interactions and on the unstructured interviews.

6.1 DISCUSSION

In general the table triggers some interest and it mainly fosters conversation about the technology itself and only indirectly it supports reflection on the actual visit. Yet, in several cases, the appearance of an image leads to a discussion about the fresco, and in particular, for those details for which the booklet does not provide enough information. Mostly, this persuasive effect of the system takes place in moments when the group is temporarily silent. The table is looked at with more attention, and stimuli for continuing the conversation are sought. When the conversation

revives again, usually the table and the stimuli subsequently displayed are ignored. This may be due to the fact that when the table itself is not the topic of the discussion, it is a relevant source of inspiration only when the stimuli are needed to overcome a moment of social lack of interaction; but when the group can independently sustain the conversation, further stimuli are not needed and the group tends to ignore the table.

This may lead to conclude that the phase (a) above is more useful that then phase (b). Yet, a possible source of bias is due to the low performance of the word spotter when the conversation is lively because of the frequent overlaps in speech of the different individuals. Therefore, many times the stimuli presented in phase (b) were not actually related to the topic of the conversation but they appeared rather randomly selected.

Another possible bias to the interpretation that the table is not effective in supporting a conversation already in place is that our system does not consider the visual attention of the individuals while proposing the stimuli. While this is less problematic when the conversation is lagging (since the individual tries to get inspiration from the table), it is more important when the conversation is lively since the social pressure tends to focus the individual attention toward the group rather than on the table. A system able to monitor the visual attention of the individuals may choose the right moment to display a stimulus and possibly make phase (b) more effective.

Regarding the use of humorous sequences of text snippets and related details, we failed to observe any interesting effect on the group. Yet this is mainly due to fact that the sequences went always unnoticed by the subjects. The witty expressions had no special characters and floated similarly to the captions. Again, it seems that the main reason is the lack of consideration about the visual attention when this strategy is applied. This is more problematic with respect to the case of the simpler visual strategies because attention must be assured from the beginning to the end of the sequence in order to get the humorous effect.

Regarding the possibility of interaction with the table, contrary to the subjects of other WoZ studies [9], in general our subjects did not expect the table to be interactive (though someone mentioned the possibility of using Microsoft Surface). Yet, when asked, they expressed a positive attitude toward the idea of using the table as an interactive kiosk and also suggested other activities, as for example accessing email and news.

7 CONCLUSIONS

We can say that the current prototype has a moderate effect on the conversation. With the current implementation, the effect is more prominent for phase (a), when the group conversation is out of topic and in particular during the moments of silence when the individuals are looking for new topics of discussion. To better explore phase (b), we need to refine the module that monitors the conversation and in particular the word spotter that has a reduced accuracy when the conversation is more lively, so as to have a more accurate assessment of the topic being discussed.

REFERENCES

[1] G. Leinhardt and K. Knutson, *Listening in on Museum Conversations.* Altamira Press, 2004.

[2] B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* Morgan-Kaufmann, 2002.

[3] C. Rocchi, O. Stock, M. Zancanaro, M. Kruppa and A. Krüger, *The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices*. Proceedings of the International Conference on Intelligent User Interfaces, IUI-2004, Island of Madeira, Portugal, 2004

[4] D. Petrelli and E. Not, 'User-centred Design of Flexible Hypermedia for a Mobile Guide: Reflections on the HyperAudio Experience' *User Modeling and User-Adapted Interaction*, vol. 15, numbers 3-4, pp. 303-338 (2005)

[5] J. Wilpon, L. Rabiner, C. Lee and E. Goldman, 'Automatic recognition of keywords in unconstrained speech using hidden Markov models', *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, pp. 1870-1878, 1990.

[6] M. Weiser, M. and J. S. Brown, 'Designing Calm Technology', *PowerGrid Journal*, v1.01, July 1996

[7] T. Matthews, A. K. Dey, J. Mankoff, S. Carter and T. Rattenbury, *A Toolkit for Managing User Attention in Peripheral Displays*. Proceedings of UIST '04, Santa Fe, NM, 2004.

[8] H. Ishii and B. Ullmer, *Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms*, Proceedings of Conference on Human Factors in Computing Systems CHI. Atlanta, GA, pp. 234-241, 1997.

[9] J. F. McCarthy, T. J. Costa and E. S. Liongosari, *UniCast, OutCast & GroupCast: Three Steps Toward Ubiquitous, Peripheral Displays*, Proceedings of the 3rd international Conference on Ubiquitous Computing, Atlanta, GA, 2001

[10] W. W. Gaver, J. Bowers, A. Boucher, H. Gellerson, S. Pennington, A. Schmidt, A. Steed, N. Villars and B. Walker, *The drift table: designing for ludic engagement*. CHI '04 Extended Abstracts on Human Factors in Computing Systems, 2004.

[11] N. A. Streitz, C. Rocker, T. Prante, D. V. Alphen, R. Stenzel and C. Magerkurth, 'Designing Smart Artifacts for Smart Environments', *Computer*, vol. 38, no. 3, pp. 41-49, 2005.

[12] J. M. DiMicco and W. Bender, *Group Reactions to Visual Feedback Tools* Proceedings of The Second International Conference on Persuasive Technology, Stanford University, CA, 2007.

[13] J. Sturm, O. Houben-van Herwijnen, A. Eyck, and J. Terken *Influencing Social Dynamics in Meetings through a Peripheral Display*, Proceedings of International Conference on Multimodal Interfaces, ICMI 2007, pp. 263-270, 2007.

[14] T.J. Kim, A. Chang, L. Holland and A.S. Pentland, *Meeting Mediator: Enhancing Group Collaboration with Sociometric Feedback*, Proceedings of ACM Conference on Computer-Human Interaction (CHI08), Florence, 2008.

[15D. Boud, R., Keogh, and D. Walker (Eds.), *Reflection: Turning Experience into Learning*. Kogan Page, 1985.

[16] G. Adomavicius, and A., Tuzhilin, 'Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions', *IEEE Transactions on Knowledge and Data Engineering* 17 (6), pp. 734–749, 2005.

[17] C. Rocchi, D. Tomasini, O. Stock and M. Zancanaro, *Fostering conversation after the museum visit: a WOZ study for a shared interface*. Proceedings of the Conference on Advanced Visual Interfaces, AVI-2008, Napoli, Italy, 2008

[18] P. Viola and M. Jones, 'Robust real-time face detection', *Computer Vision*, 57(2) pp.137–154, 2004.

[19] L. Bartram, C. Ware and T. Calvert, 'Moticons: Detection, distraction and task', *International Journal Human-Computer Studies*, 58 (5). pp. 515-545, 2005.

[20] R. Abrams and S. E. Christ, 'Motion onset captures attention: A rejoinder to Franconeri and Simons (2005)', *Perception and Psychophysics*, 68 (1), pp.114-117, 2006.

Unconscious Persuasion by Ambient Persuasive Technology: Evidence for the Effectivity of Subliminal Feedback

Jaap Ham[1], **Cees Midden**[1], **and Femke Beute**[1]

Abstract. In this paper we explore a fundamental characteristic of Ambient Persuasive Technology: Can it persuade the user without receiving the user's conscious attention? In a task consisting of 90 trials, participants had to indicate which of three household appliances uses the lowest average amount of energy. After each choice, participants in the supraliminal feedback condition received feedback about the correctness of their choice through presentation of a smiling or a sad face for 150 ms. Participants in the subliminal feedback condition received identical feedback, but the faces were presented only for 25 ms, which prohibited conscious perception of these stimuli. The final third of the participants received no feedback. In the next task, participants rated the energy consumption of all presented appliances. Results indicated that supraliminal feedback and subliminal feedback both led to more correct energy consumption ratings as compared to receiving no feedback. Implications are discussed..

1 INTRODUCTION

Since B. J. Fogg [1] introduced the concept, a multitude of research has investigated persuasive technology. For example, recent research indicates that feedback provided by an embodied robotic agent has persuasive effects on behavioral change [22]. That is, in this research people were given feedback about their energy conservation when carrying out washing tasks on a virtual washing machine [22]. Results indicated that social feedback (e.g., when the robot says "Your energy consumption is bad") leads to decreased energy consumption, and that social feedback had even stronger persuasive effects than factual feedback (directly indicating the amount of kWh).

Recently, researchers have started to investigate persuasive technology that makes use of Ambient Intelligence: The increasing pervasion of everyday live with information technology [see 3]. Computers—and thereby persuasive technology—are no longer bound to a specific location, but can be integrated unobtrusively into the environment. This allows new forms of influencing and offers some important advantages over more focal persuasive technologies. One of these advantages is the ability to deploy influence attempts at exactly the right time and place. For example, a device called WaterBot aims to reduce water consumption by tracking and displaying information about water consumption at the sink itself [2]. In line with the possibilities of Ambient Intelligence, scientists have developed new influencing concepts like implicit interaction [5], environmental persuasion [6], and ambient displays [4]. For example, ambient displays provide users with information by making it available in an environment through "subtle changes in form, movement, sound, color, smell, temperature or light" [4].

An example of an ambient display is perFrame [7], a persuasive picture frame to persuade people into proper sitting posture. More specifically, the perFrame is an unobtrusive interactive picture frame that displays a moving portrait of a close other. This portrait provides affective feedback dependent on the participants' sitting posture (e.g., the close other smiles when the participant sits correctly).

In line with Davis [8], we argue that a clear label for this new type of persuasive technology is Ambient Persuasive Technology. The goal of the current article is to improve the conceptual clarity about this form of persuasive technology. Therefore, we shall present a study that investigates one of the features of this concept.

We argue that one of the most fundamental characteristics of this kind of persuasive technology is that ambient persuasive technology is able to influence attitudes or behavior without conscious attention to that persuasive technology by the person being influenced. For example, the perFrames [7] described above should be persuasive without the necessity of the full, focal and conscious attention to the perFrame. However, in earlier research of ambient persuasive technology, the possibility of spending conscious attention to the persuasive technology always existed. For example, the described perFrames [8] may have been unobtrusive but were nevertheless clearly visible, and participants could easily focus their attention on it. In the current research, we will investigate this question: Can ambient persuasive technology persuade the user without receiving the user's conscious attention?

2 THE CURRENT RESEARCH

In the current research, we will use an embodied virtual agent to give participants feedback about their behaviour. Earlier research suggests that these social agents are able to function as persuasive technology [e.g., 22, 9, 10], or even as ambient persuasive technology [7]. More specifically, participants performed a task that consists of 90 trials. In each trial, participants were asked to make a straightforward choice: They were to indicate which of three household appliances uses the lowest amount of energy in an average family in a average week. These household appliances were chosen in such a way that this was not a very easy task. Participants were presented with sets of

[1] Dept. of Industrial Engineering and Innovation Sciences, Section Human-Technology Interaction, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. Email: j.r.c.ham@tue.nl, c.j.h.midden@tue.nl, f.beute@student.tue.nl.

three household appliances consisting of various combinations of three appliances that use little electricity, and six appliances that use more electricity. Each trial always had one correct answer. An embodied virtual agent (called "Robin SaveEnergy") was introduced to participants as someone who cares about energy consumption. After each choice, participants (except those in the control condition) received feedback about the correctness of their choice from the virtual agent. That is, some participants (in the supraliminal condition) received feedback about the correctness of their choice through presentation of a smiling or a sad face for 150 ms. Presenting the social feedback for this period of time allows participants to spend conscious attention to it. For participants in the control condition, no social feedback was presented, that is, these participants saw no smiling or sad faces. So, these participants received no information about energy consumption levels of these household appliances.

Importantly, we also devised a version of this persuasive technology in which the user cannot spend conscious attention to the social feedback, because it is simply presented for too short a duration to be consciously noticed. That is, participants in the subliminal condition received the same social feedback as those in the supraliminal condition, but for them the smiling or sad faces were presented only for 25 ms. Presenting information for very short durations is called subliminal priming [for an overview, see 11, 12]. When information is presented only very briefly, people are not able to consciously perceive it [11, see also, 13]. Research has indicated that when faces are presented for very short presentation times of for example 33 ms. [14] and less [e.g., 16, 17, 15], participants are not consciously aware of them being presented. A large literature suggests that concept activation by means of subliminal priming techniques can be quite influential [see, 13]. For example, Murphy and Zajonc [15] found that participants liked Chinese ideographs that were preceded by a subliminally presented smiling face better than the same ideographs preceded by a subliminally presented scowling face.

More recently, research showed that subliminal priming can also be used to prime goal-relevant cognitions, and that when the motive to pursue that goal is active, subliminal priming can be used as subliminal persuasion. For example, Karremans [18] showed that subliminal priming with the brand name of a drink increased people's choices for the primed brand, but only for participants who were thirsty.

In the current research, we will investigate whether persuasive technology that provides people with interactive feedback about their choices can influence people's attitudes without the need for conscious attention to the feedback. We argue that this research contains two innovations. First, as described, we will assess one of the most fundamental characteristics of ambient persuasive technology. Second, we will assess a completely new usage of subliminal priming. That is, earlier studies of subliminal priming and subliminal persuasion have all presented people with fixed information [e.g., a brand name, 18]. The current research (to our knowledge) will be the first to investigate whether subliminal priming can successfully be used to give interactive feedback.

After 90 of trials of choice and feedback (supraliminal, subliminal or none), we assessed participant's attitudes by asking them to rate the energy consumption of all presented appliances. Considering that ambient persuasive technology is able to influence attitudes without the need for conscious attention to the persuasive technology, we expected not only participants who received supraliminal feedback but also those who received subliminal feedback to rate the energy consumption levels of these appliances more in line with their actual energy consumption levels than participants who received no feedback.

2.1 Method

2.11 Participants and design

Sixty-one participants (39 men and 22 women) were randomly assigned to one of three experimental conditions: a supraliminal feedback condition, a subliminal feedback condition, and no feedback condition. All participants were native Dutch speakers. The experiments lasted 25 minutes, for which participants were paid 5 Euros (approximately $6.75 U.S. at the time this study was conducted).

2.11 Materials and procedure

Participants were invited to engage in an experiment ostensibly testing their knowledge about household appliances. Upon arrival, they were seated individually in a small room in front of a computer. Next, participants in all three conditions were asked to perform a first task. The purpose of this task was to make the participants familiar with nine household appliances, each depicted on a different picture. In this task, the picture of a household appliance was presented on screen, and participants were asked to indicate how often they used (per week) this type of appliance. Participants could answer by choosing a category (ranging from 'several times a day' to 'less than once a week'). Trials were presented in random order.

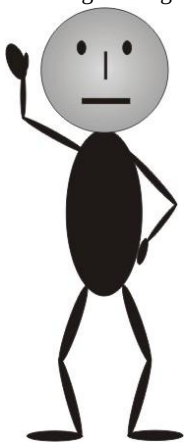

Figure 1. Robin SaveEnergy

After this task, participants were asked to conduct a second task. This task started with the introduction of a drawn figure (see Figure 1). This drawn figure was introduced as 'Robin SaveEnergy', and participants were told that this manikin had a strong opinion about saving energy. That is, they were told that for Robin SaveEnergy it is very important to save energy. Also, they were told that to be able to save energy, it is vital to have knowledge on the energy use of a household appliance during an average week in an average household, and that Robin has this knowledge. Participants were told that the task (choice task) they were about to begin with would measure their knowledge on the average energy use of a certain appliance in an average household per week. After this introduction, the task was explained in detail. Participants were told that on each screen of this task, they would be presented with the pictures of three household appliances, and their task was to indicate which one used the lowest amount of energy in an average household in an average week. More specifically, participants were asked to press a key on the keyboard corresponding to their choice as quickly as possible. In addition, they were told that Robin SaveEnergy would be

Figure 2. A screenshot of the choice-task

watching their performance in this task. Participants were told that after making their choice, a dot would appear for half a second above the three appliances, and that the face of Robin SaveEnergy would appear next at that same location. They were instructed to press the spacebar as quickly as possible when Robin's face appeared to acknowledge that they had seen him. Participants first completed five practice trials, and then the 90 trials of the experimental task started.

Each trial (of the practice task and the experimental task) consisted of a screen on which three pictures of three different household appliances were presented (see Figure 2). Participants could indicate the one that used the least amount of energy by pressing either the '1', '2', or '3' key. After one of these keys was pressed, a (focus) dot was presented for 500 ms above the middle one of the three pictures. More specifically, participants in the supraliminal feedback condition were given feedback by presenting a picture showing the smiling face of Robin (see Figure 3) was presented in case of a correct answer, or a picture showing the sad face of Robin (see Figure 4) was presented in case of an incorrect answer. In both cases, the face was presented for 150 ms. For participants in the subliminal feedback condition, the same faces were presented, but only for 25 ms. In all conditions, feedback (a smiling or sad face) was preceded by a premask and immediately followed by a postmask. These masks were used to tightly control the duration of the presentation of the faces (presented for 22 ms), and consisted of a square (equal in size to the faces) filled with random dots. Each mask was presented for 110 ms (for a detailed discussion of subliminal priming methods, see [11]). For participants in the no feedback condition, the procedure of these trials was identical (including presentation of the masks), but no feedback was given, that is, neither smiling faces nor sad faces were presented. Finally, for participants in all conditions the neutral face of Robin (see Figure 3) was presented until they pressed the spacebar. Premask, feedback, postmask, and the neutral face of Robin were all presented at the same location as the (focus) dot.

Figure 3. Happy face

Figure 2. Sad face

Figure 5. Neutral face

Previous research using similar subliminal presentation procedures has demonstrated that participants are unable to consciously perceive a presentation of 50 ms or less [e.g. 18, 23, 24]. To check whether our participants had consciously perceived the presentation of sad or smiling faces, participants in the supraliminal and subliminal feedback conditions were debriefed and checked for awareness of these faces using a funnelled debriefing procedure. Results of this debriefing indicated that all participants in the subliminal feedback condition were unaware of these sad or smiling faces, whereas all participants in the supraliminal feedback condition reported having seen these faces.

To describe the nine appliances, we used pictures of a microwave oven, a water cooker, a flatiron, a coffee making machine, a computer, an audio set, a washing machine, a vacuum cleaner, and a television set. The first three of these are household appliances that use little energy in an average family in an average week, whereas the last six use more (www.milieucentraal.nl). In every trial (of the practice task and the experimental task), one of three low-energy-consumption appliances was presented in any possible combination with two of the six high-energy-consumption appliances. Furthermore, the three low-energy-consumption appliances were never displayed together. For each low-energy-consumption appliance, 15 combinations with the other appliances were possible. The total of 45 (3 times 15) combinations was presented twice to form the 90 trials of the experimental task. For the five practice trials, five random selections of possible combinations between five different household appliances were presented.

Next, all participants answered the questions that served as the dependent measures. More specifically, participants were asked to rate the energy consumption levels of all nine household appliances. On nine different screens, the picture of a household appliance was presented together with the question "How much energy does this appliance use in an average household in an average week?" and participants could indicate their answer on a 5-point rating scale (1=very low energy consumption, 5=very high energy consumption). So, three of these questions were about the three household appliances that used little energy, and the mean answer to these three questions was our first dependent variable. The mean ratings of the energy consumption of the six appliances that use more energy functioned as the second DV.

Finally, participants answered several demographic questions, were debriefed and thanked for their participation.

2.1 Results

The average energy consumption ratings for the low energy consumption appliances and the average energy consumption ratings for the high energy consumption appliances were submitted to a 3 (feedback condition: supraliminal feedback vs. subliminal feedback vs. no feedback) x 2 (appliance type: low energy consumption vs. high energy consumption) MANOVA, with the last variable within-subjects. This analysis showed that all participants correctly rated the low energy consumption appliances as consuming less energy (M = 2.7, SD = .7) than the high energy consumption appliances (M = 3.6, SD = .5), $F(1, 58) = 51.54$, $p < .001$, $\eta^2 = .47$. However, this effect was qualified by condition, that is, we found a significant interaction of Appliance Type x Feedback Condition, $F(2, 58) = 4.39$, $p = .017$, $\eta^2 = .13$. As expected, special contrast analyses indicated that participants who had received *supraliminal* feedback indicated a bigger

difference in energy consumption ratings between the low energy consumption appliances and the high energy consumption appliances (M = 2.6, SD = .5 vs. M = 3.8, SD = .4) as compared to participants who had received no feedback (M = 3.0, SD = .8 vs. M = 3.3, SD = .5), $F(1, 58) = 8.24$, $p = .006$, $\eta^2 = .12$. Intriguingly, as expected, participants who had received *subliminal* feedback also indicated a bigger difference in energy consumption ratings between the low energy consumption appliances and the high energy consumption appliances (M = 2.6, SD = .7 vs. M = 3.6, SD = .5) as compared to participants who had received no feedback, $F(1, 58) = 4.52$, $p = .038$, $\eta^2 = .07$. Finally, a direct comparison did not suggest that participants who had received *subliminal* feedback indicated a bigger difference in energy consumption ratings between the low energy consumption appliances and the high energy consumption appliances as compared to participants who had received *supraliminal* feedback, $F < 1$. All means and standard deviations are presented in Table 1.

Table 1 Mean Energy Consumption Rating (and Standard Deviations) for Low Energy Consumption Appliances and High Energy Consumption Appliances by Feedback Type

	Feedback Type		
Appliance Type	Supraliminal	Subliminal	No Feedback
Low Energy Consumption Appliances	2.6 (.5)	2.6 (.7)	3.0 (.8)
High Energy Consumption Appliances	3.8 (.4)	3.6 (.5)	3.3 (.5)

Note. Standard deviations between brackets.

3 GENERAL DISCUSSION

Results of the current study indicate that not only participants who received supraliminal feedback but also those who received subliminal feedback gave more correct ratings of the energy consumption levels of household appliances than participants who received no feedback. Remarkably, feedback that was presented subliminally--only for 25 ms.--led to an influence on energy consumption ratings that was similar to the one caused by supraliminal feedback that was presented for 150 ms. and could clearly receive conscious attention and processing. This subliminal feedback was presented too short to receive conscious attention [see e.g., 14], nor did any of the participants in this condition indicate having (consciously) seen the smiling faces or the sad faces. The current research is in line with earlier work that shows that subliminally presented faces can influence attitudes or behaviour [e.g., 14, 15, 16, 17].

So, can persuasive technology persuade the user without receiving the user's conscious attention? The current results suggest that it can. In addition, these results suggest that (at least in the current task) interactive feedback that people can spend attention to has the same influence as interactive feedback that people cannot spend attention to. Of course, future research should further investigate this first and intriguing finding. But the current results do suggest that ambient persuasive technology of which people are not consciously aware may have an influence on people's attitudes, and that influence could (under certain conditions) be comparable to the influence of persuasive technology that needs focal attention.

We argue that persuasive messages that people do not have to spend conscious attention to can have various advantages. That is, people do not have to spend cognitive effort on it, and can still process it when low on cognitive resources. For example, the current results suggests that the perFrames described earlier [7] might have an influence on sitting posture even when people do not spend conscious attention to this interactive picture frame (e.g., after it has been on their desk for 6 months). Also, people may not become annoyed by the persuasive attempts of ambient persuasive technology—at least there is no focal influencing technology to become annoyed about, and the same might hold for becoming reactant [see, 19] towards influence attempts.

We argue that there are at least two possible social cognitive mechanisms for the influence of subliminal feedback in the current study. First, the subliminal feedback given to participants in the current study may have exerted its influence through *subliminal evaluative conditioning* [see e.g., 12]. In evaluative conditioning, the pairing of presentation of an object with presentation of a negatively or positively valenced stimulus will eventually lead to the acquisition by that object of the same positive or negative experienced value. An example of subliminal evaluative conditioning is research by Krosnick [20] that demonstrated that the evaluation of a target person can be influenced by repeatedly pairing photographs of that target person with positively or negatively evaluated events. So, the influence of subliminal feedback in the current research may have occurred through subliminal evaluative conditioning [see also, 15]. By pairing the household appliances that use little energy with smiling faces, and the other household appliances with sad faces, the general evaluation of these appliances may have been conditioned, and that may have influenced the energy consumption ratings made by our participants. Second, the subliminal feedback given to participants may have exerted its influence through goal-striving related processes [see e.g., 21]. That is, earlier research indicated that non-conscious goal pursuit can occur. For example, Hassin [21] primed participants with the goal to be flexible without making them aware that this goal had been activated. In a next task, without knowing why, participants showed to be more open minded towards other people. In the current research, comparable non-conscious goal-striving processes may have been at work. But in contrast to earlier research, people were aware of activation of the goal, but unaware of being (subliminally) primed with the feedback information that they needed to attain their goal.

A remaining issue concerns the social nature of the feedback. Our study cannot distinguish between the evaluative and the social nature of the feedback. Recent studies by Vossen, Ham and Midden [25] and Midden and Ham [26] demonstrated that both the evaluative and the social nature of feedback could add to the total feedback effect. This evokes for future work the question whether both effects would occur as well at the subliminal level.

Finally, we like to point at potential ethical issues of subliminal information. Interventions that go beyond the control of the receiver should be regulated carefully. The most important reason for this is that subliminal priming manipulations clearly lack in transparency. In particular, this will be necessary for applications that are beneficial for the sender. Informed consent by the receiver seems a crucial aspect of subliminal persuasion.

Obviously, ethical issues should be an issue for further consideration.

This research sheds light on a fundamental characteristic of ambient persuasive technology: It is able to influence attitudes *without conscious attention* to that persuasive technology by the person being influenced. Future research could investigate whether also behaviour can be influenced this way. With this work, we want to help improve the conceptual clarity about this form of persuasive technology, and we have laid out a methodology to investigate related issues.

4 ACKNOWLEDGEMENTS

We thank the Persuasive Technology Lab Group at TUe and Johan Karremans for their comments on this research project.

REFERENCES

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

[2] E. Arroyo, L. Bonanni, and T. Selker. Waterbot: Exploring feedback and persuasive techniques at the sink. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005), pages 631–639, 2005.

[3] Riva, G., Vatalaro, F., Davide, F., Alcaniz, M. (eds.) Ambient Intelligence. IOS Press, Amsterdam (2005)

[4] Wisneski, C., Ishii, H., Dahley, A., Gorbet, M.G., Brave, S., Ullmer, B., Yarin, P.: Ambient Displays. In: Streitz, N.A., Konomi, S., Burkhardt, H.-J. (eds.) CoBuild 1998. LNCS, vol. 1370. Springer, Heidelberg (1998)

[5] Schmidt, A.: Interactive Context-Aware Systems Interacting with Ambient Intelligence. In: Riva, G., Vatalaro, F., Davide, F., Alcaniz, M. (eds.) Ambient Intelligence. IOS Press, Amsterdam (2005)

[6] Mathew, A. P.: Using the environment as an interactive interface to motivate positive behavior change in a subway station. In CHI '05 Extended Abstracts on Human Factors in Computing Systems, pages 1637–1640, 2005.

[7] Obermair, C., Reitenberger, W., Meschtscherjakov, A., Lankes, M., & Tscheligi, M.: perFrames: Persuasive picture frames for proper posture. In: Oinas-Kukkonen, H. et al., (Eds). Persuasive 2008, LNCS, vol. 5033, pages 128-139, Springer, Heidelberg (2008).

[8] Davis, J.: Towards participatory design of ambient persuasive technology. Presented at Persuasive Pervasive Technology and Environmental Sustainability, Workshop at the Sixth International Conference on Pervasive Computing (Pervasive 2008), Sydney, May 19–22, 2008.

[9] Grolleman, J., van Dijk, B., Nijholt, A., van Emst, A.: Break the Habit! Designing an e-Therapy Intervention Using a Virtual Coach in Aid of Smoking Cessation. In: IJsselsteijn, W., de Kort, Y., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 133–141. Springer, Heidelberg (2006)

[10] Creed, C.: Using Computational Agents to Motivate Diet Change. In: IJsselsteijn,W., de Kort, Y., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 100–103. Springer, Heidelberg (2006)

[11] Bargh, J. A., & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In H. Reis & C. Judd (Eds.), Handbook of research methods in social and personality psychology (pp. 253–285). New York: Cambridge University Press.

[12] Dijksterhuis, A., Aarts, H., & Smith, P. K. (2005). The power of the subliminal: Subliminal perception and possible applications. In R. Hassin, J. Uleman, & J. A. Bargh (Eds.), The new unconsciousness (pp. 77–106). New York, NY: Oxford University Press.

[13] Strahan, E. J., Spencer, S. J., & Zanna, M. P. (2002). Subliminal priming and persuasion: Striking while the iron is hot. Journal of Experimental Social Psychology, 38, 556–568.

[14] Hsu, S-M., Hetrick, W. P., & Pessoa L.: Depth of facial expression processing depends on stimulus visibility: Behavioral and electrophysiological evidence of priming effects. Cognitive Affective and Behavioral Neuroscience 2008, Sep, 8, 282-92.

[15] Murphy, S. T., & Zajonc, R. B. 1993. Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposure. Journal of Personality & Social Psychology, 64, 723-739.

[16] Niedenthal, P. M. 1990. Implicit perception of affective information. Journal of Experimental Social Psychology, 26, 505-527.

[17] Ric, F. 2004. Effects of the activation of affective information on stereotyping : When sadness increases stereotype use. Personality and Social Psychology Bulletin, 30, 1310-1321.

[18] Karremans, J. C., Stroebe, W., & Claus, J. 2006. Beyond Vicary's fantasies: The impact of subliminal priming and brand choice. Journal of Experimental and Social Psychology, 42, 792-798.

[19] Brehm, J. W. 1966. A theory of psychological reactance. New York: Academic Press.

[20] Krosnick, J. A., Betz, A. L., Jussim, L. J., & Lynn, A. R. 1992. Subliminal conditioning of attitudes. Personality and Social Psychology Bulletin, 18, 152–162.

[21] Hassin, R. R. (2008). Being open minded without knowing why: Evidence from nonconscious goal pursuit. *Social Cognition, 26*, 578-592.

[22] Midden, C., & Ham, J. 2008. The persuasive effects of positive and negative social feedback from an embodied agent on energy conservation behavior. Paper presented at AISB 2008, Aberdeen, Scotland.

[23] Fishbach, A., Friedman, R. S., & Kruglanski, A. W. 2003. Leading us not onto temptation: Momentary allurement elicit overriding goal activation. *Journal of Personality and Social Psychology, 84,* 296–309.

[24] Shah, J. Y., & Kruglanski, A. W. 2002. Priming against your will: How accessible alternatives aVect goal pursuit. *Journal of Experimental Social Psychology, 38*, 368–383.

[25] Vossen, S., Ham, J. R. C., & Midden, C. J. H. 2009. Social influence of a persuasive agent: The role of agent embodiment and evaluative feedback. Manuscript submitted for publication.

[26] Midden, C. J. H., & Ham, J. R. C. 2009. Using negative and positive social feedback from a robotic agent to save energy. Manuscript submitted for publication.

Social networking sites as platforms to persuade behaviour change in domestic energy consumption

Derek Foster[1], Mark Blythe[1], Shaun Lawson[2] and Mark Doughty[2]

Abstract. This paper describes a pilot investigation into the use of the social networking site Facebook as a platform for persuasive applications. The application domain is behaviour change in domestic energy consumption and the study focuses on determining peoples' attitudes towards the hypothetical coupling of the consumer product Wattson, which can monitor domestic electricity usage, to a Facebook application termed Watts Up. The Facebook application presents visualisations of users' own electricity consumption as well as that of their friends. Users' attitudes towards this notion were accumulated and analysed using grounded theory. Some user indications revealed negative opinions about the concept based, for instance, around privacy and confusion; however the balance of opinion appeared to favour the underlying idea that revealing other people's energy usage data would lead to competition and peer influence to reduce energy consumption.

1 INTRODUCTION

The responsible consumption of energy in domestic homes is a topic of considerable current importance. Heightened awareness of the global negative impact of the depletion of non-renewable energy resources and global warming, coupled with both the current economic situation and rising energy bills, means that many household consumers are increasingly likely to be motivated to reduce their energy consumption. However, research [1] has shown that many people are unaware of the day-to-day cost of, for instance, electricity consumption and that the obfuscation of information presented on household meters and quarterly bills presents a significant barrier to the understanding of daily energy use.

So-called *smart-meters* which present information in ways that are more easily understood by the consumer (e.g. via visualisations of daily/weekly usage presented online in web-pages) have recently become a focus of research and evaluation (e.g. see [2]); however – at least in the UK – such devices are not yet routinely made available through energy suppliers. Nonetheless, individuals can choose to purchase one of a range of consumer devices that will show, in a real-time fashion, their own electricity usage to them within their home. These devices include the Owl [3] and Eco-Eye [4] which both display energy usage in a relatively utilitarian fashion. An alternative product is the Wattson [5] by DIY Kyoto which embraces a more aesthetic design ideal (its developers include alumni of London's Royal College of Art) and could be categorized as an ambient display [6]. The Wattson is shown in a domestic setting in Figure 1. The device can show instantaneous usage of electricity in kWatts or as an annual fee in UK £pounds. It also has a series of LEDs embedded in its base which emit an ambient colour based on current usage (such as blue for low usage). Users of the Wattson can optionally choose to use a piece of software which logs their usage and presents it via visualisations. Furthermore, users can also choose to upload their usage statistics to the Kyoto website and make it publicly available to other users. Although the take-up of this functionality seems limited (based on casual observation of forums on the DIY Kyoto website) it is a step towards the persuasive use of such technology as it promotes awareness of other people's usage as well as one's own.

Figure 1. Wattson Device in a home setting

Social networking sites such as Facebook and MySpace have seen phenomenal growth in the number of people using them in a very short space of time. Reportedly, Facebook has in excess of 200 million current users [7]. The emergence of freely available software development tools such as Facebook Platform and OpenSocial has released the potential of deploying small software applications to very large numbers of people in a viral fashion. The success of seemingly trivial applications on Facebook has shown that people are willing to invest daily time in interacting with the applications they install as well as recommending such applications to their friends. The Facebook application metrics site AppData [8] shows top installs having many millions of monthly users whilst even the most trivial of applications can have tens of thousands of users. Given the number of users and their social connectivity – it seems logical that Facebook could provide a very powerful platform for the delivery of persuasive applications and this has indeed been suggested by a number of researchers (e.g. [9] [10]). Nevertheless, the number of applications that have been

[1] Dept. of Computing Science, Univ. of York, YO10 5DD, UK. Email: **derek@derekfoster.net , mblythe@cs.york.ac.uk**

[2] Lincoln Social Computing Research Centre (LiSC), Dept. of Computing & Informatics, Univ. of Lincoln, LN6 7TS, UK. Email: {**slawson**, **mdoughty**}**@lincoln.ac.uk**.

evaluated in an academic context for this purpose remain very limited.

Figure 2. User's view of their hypothetical energy usage once they install Watts Up in their Facebook profile.

Social Psychology can offer us an insight into why social platforms like Facebook can be powerful motivators into behaviour change. Individuals join Facebook *voluntarily* and add applications to their profile on the same basis. The main function of Facebook is to provide the user with an online network of their chosen friends they can interact with in various ways. The *attractiveness* of such friends in terms of their similarity and familiarity to us makes it more appealing for social interaction to take place [11]. This forms the basis of interaction between friends and a primary reason why we add them to our 'friends list'. There is likely to be more attitudinal change between friends when the friendship attributes of familiarity and attractiveness of other friend's qualities are present.

Very recent investigations of Facebook have produced a series of named patterns that attempt to spread persuasive behaviour by embedding them in applications [12]. These patterns make use of the built in features of the platform such as the friend selector and messaging functions to mitigate the spreading of an application through a social network in a viral fashion. One such pattern is called 'Provoke and Retaliate' where one friend can take action on another friend, for example by sending a virtual gift or a graphical representation of encouragement. This generates reciprocity whereupon the friend on the receiving end feels socially obligated to respond. Using this concept it can be supposed that reciprocal interaction could take place in an energy monitoring application where friends view each other's energy information and can both send/receive encouragement and warnings based on their energy use. Other persuasive factors are Cognitive Dissonance [13] and Group Polarization [14].

Cognitive dissonance is manifested when someone holds two or more inconsistent beliefs; an example of a person's energy awareness attitude could comprise, for instance two statements such as (i) I know that energy consumption is having a bad impact on the environment, but also (ii) I always leave the lights and heating even when not at home. These conflicting ideas can induce cognitive dissonance therefore creating more awareness and in turn a drive to reduce the dissonance through changing their behaviour and attitude. In this case the behaviour change would be to stop leaving the lights and heating on to align with their belief of energy consumption having a negative impact.

Group Polarization was a term that was known as' Risky Shift' up until the 1960's. Risky Shift was the idea that an individual makes more risky decisions than a group. However further research proved this to be incorrect, with the realization that groups make more 'extreme' decisions than individuals [15]. More recent research has revealed that when Group Polarization happens online its effect is even greater [16], therefore increasing the potential persuasiveness of an application in an online setting. This effect shows that if a group of people have a meeting and discuss a particular topic such as Climate Change then the participant's ideas and views are strengthened even more than they were before, since such group meetings create a highly persuasive environment. However this can also have a negative effect: for example, if an inherently racist group of people have a discussion about racism then their views on racism are inflated and strengthened. Group Polarization is manifested in groups of Facebook friends who are part of a particular group with a common theme such as 'Feed a child with a click' [17] that raises awareness of children dying from starvation, thereby strengthening the views of the users who join the group. Such groups can be created for energy awareness and discussion areas integrated into applications such as Watts Up. By doing so, Group Polarization could be leveraged creating a persuasive environment where a group of Facebook friends share common beliefs based around an energy conservation theme.

In this paper we present an investigation of people's attitudes to the energy consumption monitoring device Wattson and also to a hypothetical software application that allows people to view their own and their friend's electricity usage via a Facebook application. The paper is organized as follows. Firstly, in Section 2, we describe the Facebook application and its hypothetical link to the Wattson. Next, in Section 3, we describe our user evaluation of such a concept interpreted using grounded theory. Section 4 presents the analysis of the collected data interpreted using grounded theory. Finally, section 5 presents our conclusions and provides brief directions for future work.

2 RELATING WATTSON TO FACEBOOK

The focus of this paper is on the hypothetical use of the Wattson device in combination with a conjectured Facebook application named 'Watts Up'. The concept explored is the capability of the Wattson to send measurements of the user's current energy usage to the Watts Up application. When the user visits Watts Up, they can see their own energy usage represented numerically and graphically from within their Facebook account. Additionally, they can also check on any of their friend's energy usage if they too use the Wattson device with Watts Up. It is envisaged that by being able to view other friend's energy ratings, it may introduce a 'competitive' element where users thereby compete by lowering their energy usage, thus having a positive impact on the environment. Social Approval [18] would be a key element as a user who consistently used more energy than the rest of their friends would likely be liable to provocation by their friends in the form of comments or any other persuasive facilitators the application offers. By 'conforming' to the norm the user's mental model of acceptance by their friends is

consolidated, particularly through reciprocal positive encouragement. It should be noted that the Wattson is not actually used throughout the investigation; rather it is a simulation of its potential use when coupled with the Watts Up application. The user interface to the Watts Up application however was implemented using the Facebook Platform and made available as an application to users on Facebook.

The concept of Social Learning has shown that people can learn new behaviours and adopt attitudes by observing. Through this observation they can note possible rewards and punishments and are likely to perform this rewarding behaviour themselves [19]. As such, a form of league table or virtual reward scheme could be implemented in 'Watts Up' to facilitate the notion of good energy conservation practice being rewarded, particularly when those rewards are not only automatically granted by the system, but also by their friends. This can be reinforced further particularly if the friend giving a reward is older or more experienced. Examples of this are community websites that encourage people to stop bad habits, such as smoking [20], [21]. These sites post up success stories of people who have quit smoking. Additionally, they also provide the facilities to send encouragement messages to others that can be personal or posted publicly on a board where others can take inspiration and encouragement from them.

Shown in figure 2 is the applications tab 'My Watts' which illustrates the user's current and previous energy usage both numerically and graphically. Visualisations incorporating the typically emotive polar bear image have also appeared in related work by other authors [22]. Another important function of the application is the 'Friend Watts' tab adjacent to the 'My Watts' tab, illustrated in figure 3. This shows a list of the user's friends who have added the application and also shows in brief a graphical representation of their overall energy usage. The decision was made during development, not to include or show the energy usage of others who use the application who are not on your friends list. This can give rise to the 'Fake Friends' syndrome that Facebook is trying to avoid. The reasoning behind this is that people may add these friends to their friends list even though they do not know them at all. The consequence is Facebook's goal of an accurate social graph being completely wrong. Facebook themselves have tried to address this issue partly in their launch of the new profile layout in July 2008 which moves applications of profile pages and onto boxes possibly making them more difficult to find.

3 USER STUDY

The purpose of the short study was to determine peoples' attitudes towards the coupling of the Wattson to the hypothetical Facebook application Watts Up. A number of participants, 10, were recruited – using Facebook messaging – to take part in the study, 7 participants completed the questionnaire. The method used to gather the qualitative data necessary for analysis of potential user thoughts and opinions was a questionnaire utilising open-ended queries [23]. Once gathered the data was analysed using the Grounded Theory (GT) method [24]. The questions asked were designed to encourage users to challenge the design aspects of both the Wattson and the Watts Up Facebook application in order to reveal potential good points as well as the bad points of their respective designs and functionality. The intention was to provide viewpoints for possible re-design elements to potentially improve the concept of energy awareness in the home.

Figure 3. User's view of their friends' hypothetical energy usage in the Watts Up application.

The questionnaires used in the study were embedded into the Watts Up application, with the user being able to complete and submit their answers from within their Facebook account. This provided a more integrated solution for capturing the data as the user will have viewed the features in Watts Up directly before commencing the questionnaire. A screenshot of the questionnaire embedded within Watts Up and the users Facebook account is seen in figure 4.

Figure 4. Questionnaire embedded in Facebook Application.

Two sets of questions were used in the study – one evaluating the Wattson device itself and the other exploring the notion of linking the Wattson (or similar) to a Facebook application. The questions posed to users in order to evaluate their thoughts and attitudes on the potential use of the Wattson device are detailed

in Table 1. Additionally, the questions posed to users in order to evaluate the Facebook application are given in Table 2.

Question	Phrasing
1	From your own thoughts regarding what you have learned so far about the Wattson device, describe how you think it may be able to assist you to personally contribute towards energy conservation. For example it may help to assist in achieving a cheaper home electricity bill.
2	What do you think about the Wattson's design in terms of its physical appearance? For example could it complement your home furnishings?
3	What improvements do you think would benefit the device in terms of its physical appearance? For example its size.
4	From your understanding of the device so far, please describe any barriers to using it effectively. For example is your fuse box located outside your home.
5	Would you recommend the device to your friends and family? Please give reasons for your answer.

Table 1. Questions used to evaluate the Wattson device

Question	Phrasing
1	What are your thoughts on this Facebook Application (Watts Up) in how it complements the Wattson device? For example, how do you feel it would be a useful addition to it?
2	Please describe your experience when using the Watts Up application. For example were its features easy to understand or not?
3	When using the Watts Up application, please describe what your own interpretation of your energy rating was. An example could be your thoughts on the illustrative representations (i.e. the polar bear) of your energy rating
4	In the applications 'Friends Watts' feature, please describe whether you feel this feature might be able to make you more aware of your own energy usage at home by comparing with friends. Please be as descriptive as possible.
5	Would you recommend this application if it were to be fully developed? Please explain your reasons why.

Table 2. Questions used to evaluate Watts Up Facebook Application

Other related studies that have taken place to raise awareness of domestic energy consumption have involved the use of persuasive computer games [25]. This involves a type of game play design where users can interact with a simulated domestic home environment. Within this they can perform various energy-usage actions such as taking a shower, watching television or cooking a meal. Instant feedback of energy usage is given on screen as well as a monetary meter which drops as they use more energy. It allows for a cause and effect simulation to take place with instant graphical feedback which isn't normally available when carrying out day to day activities. A pre-study of 100 teenagers aged 13-18 took place to determine how important they found energy awareness to be, with 70% deeming it was important. The authors have suggested that subsequent empirical research using their game design will provide a greater insight into energy conservation in the home.

4 ANALYSIS OF USER COMMENTS

In this section the data collected using the questionnaire is presented and analysed. A total of 7 questionnaires were completed which produced approximately 3,300 words of qualitative data, as derived from the responses to the questionnaire's open ended questions. Each questionnaire included responses to both sets of questions for the Wattson and for the Watts Up application. To analyse the data, as stated previously the chosen qualitative research method was GT. The format of the questionnaire was designed to encourage and elicit from the participants both the good and bad attributes of the Wattson device and the Watts Up application. By using the GT method in its procedural steps on the questionnaire data, the intention was to present a theory that encapsulates views to re-designing aspects of the aforementioned device and application. The intention therefore is that the theory viewpoints will be entirely grounded in and from the data. The GT method was applied to both the Wattson device and 'Watts Up' Facebook application evaluations separately, with reporting carried out on both. Due to space restrictions we will only fully report and elaborate on the Watts Up application, mainly as this is the level at which most of the social interaction takes place. The various steps of applying the GT method to the data will now be detailed: these are open coding, axial coding and selective coding.

Sample of user comments listed in the open coding stage, as well as their generated codes are given in Table 3 for the Watts Up application.

User Comment	Codes generated
"I think it would be useful as an anywhere access insight into your home energy needs, but perhaps more from a home automation perspective. I confess I would use it, but think it could be improved to say track usage over time."	1. Implies **anywhere access** is **useful** as a **feature** 2. Implies that **improvements** to the **presented data** could be made
"for looking back at your usage to perhaps see if there is a pattern in when it has been high so that you can change behaviors'	1. **Energy history** seems to be important 2. Implies that you will **lower energy** usage when it becomes high
"non technical minded people can use it"	1. Implies that **non-technical** people can use it easily
"Not completely sure how this application works with the device", , "I would probably not regularly check my friends energy usage."	1. Implies confusion in how the application **communicates** with the device 2. Implies **friends feature** isn't interesting
"score between my friends will encourage us to improve our usage"	1. Implies **encouragement** from friends will improve usage
"fear of having someone know how much energy I use."	1. Implies **fear** or **embarrassment** on **friends** knowing energy use
"beat their friends"	1. Implies **competitiveness** element to beat friends in usage
"found it quite simple."	1. Implies the app had a **simple layout**
"tabs make it easy to navigate"	1. Implies tabs provided **good navigation**
"I did not find the score output method obvious."	1. Implies the **energy rating** was **unclear**

Table 3. User comments and open coding for Watts Up

Following open coding, the axial coding step was then performed in order to determine commonalities between the codes. Three categories were drawn out at this step: *usability*, *engagement* and *confusion*. These categories are shown as headings in Table 3 which also shows the grouping of the relevant codes underneath these headings. Figure 5 shows a simplistic diagram illustrating the relationships between the three

axial codes. The *usability* category was defined since many of the codes derived from the data were specifically related to the application layout and the presentation of its features and functions. For example the 'anywhere access' label was deemed important as the application offered the capability to view your homes energy usage from any internet browser. This could be projected as greatly enhancing the functionality of the Wattson device, as without the application its measurements could only be viewed by the user when present in their home. The *engagement* category was directly influenced by the application's usability credence. As shown by the codes under engagement, there is a definite mixture of interaction taking place. A whole range of emotions are instigated: from fear of friends mocking another friend's energy usage (leading to embarrassment) through to constructive and friendly competitiveness. Peer pressure was noted in the feedback as being an important construct that can spur people on to contribute to energy conservation - but it was also suggested that this may also have a potentially negative impact.

Usability	**Engagement**	**Confusion**
simple layout	competitiveness	unhelpful
non-technical	useful	reluctance
energy history	fear	visuals meaning
anywhere access	baseline rating	poor comparison
energy rating	peer pressure	help key
representation	encouragement	unclear
communicates	challenge	complacency
enhances	feel guilty	improvements
numerical/graphical icons	environment awareness	
	friends	
	embarrassment	
	comparing	

Table 4. Categories drawn out by axial coding of user feedback

There was also an element of *confusion* which highlights a failing in areas of the application's usability and engagement. There was a relatively moderate degree of confusion surrounding the use of the current graphical representations such as the polar bear and earth light bulb. However, there were also several unique suggestions in the data to combat this, such as utilisation of a help indicator or key to immediately identify what the graphical illustrations represent. This not only states a possible solution but also gives the impression from the feedback that individuals have a deeper understanding of the environmental impact by offering up alternative solutions.

Following axial coding, GT advocates the determination of a *core category* – in this case this was deemed to be *engagement*. This was chosen because there was a predominate amount of user data describing a strong link between the human interaction with the application to produce an effect on the environment. There was a strong sense of social interaction (with friends) within the data that is related across all the categories of usability, engagement and confusion. It was found that the application, although limited in interaction between the user and the application interface, still provided a major link to external social aspects through friends linked by the application. These social aspects were manifested as human behaviour between friends such as competitiveness, peer pressure and encouragement. The aforementioned behaviours were positive in general but negative emotions were also used: such as guilt and embarrassment when a user's energy usage was higher than their friends. From this it can be said that the application could be a strong motivator to incite these behaviours and emotions with the environment the chief benefactor. Additionally the application could provide a 'baseline' for a user's energy usage by comparing to their friends energy usage. However, some users would prefer not to interact with the friends feature at all giving rise to redevelopment. With some confusion on the meaning of the graphical representations for a personal energy rating the use of well placed help information could assist greatly.

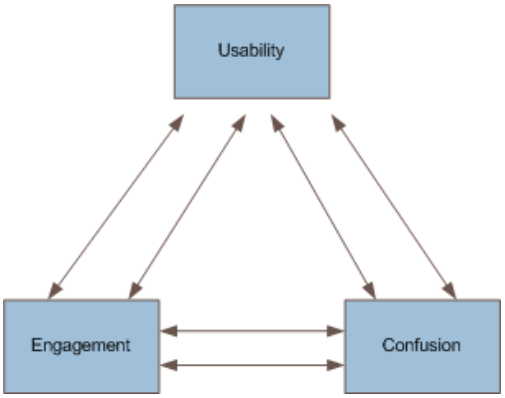

Figure 5. Axial coding category relationships

It is apparent that we can deduce a number of potential redesign concepts for the Watts Up application. For instance, the graphical representation (polar bear, earth light bulb) could be changed to give a more understandable energy rating, in terms of redesigning it would basically be a new set of images. Help icons could also be situated strategically to assist users with the meaning of anything on applications pages, would be helpful as raised by the feedback regarding confusion of graphical representations. The potential to give an alphanumeric overall rating to a user's home could be introduced such as the current rating scheme in use for cookers, fridges, washing machines etc. This could be integrated on the user 'My Watts' page of the application. Finally, to combat privacy, the option to completely disconnect from other users of the application could be integrated, effectively leaving the user isolated with their energy data remaining private.

5 CONCLUSIONS & FUTURE WORK

The paper has described a short pilot investigation into user attitudes towards coupling home energy monitoring devices, such as the Wattson, with Facebook applications which reveal not only individual energy usage data but that of people's friends. Analysis of user statements, based on Grounded theory, revealed a number of categories that must be considered in any future real deployment of such a concept. Some user indications revealed negative thoughts about the concept based, for instance, around privacy and confusion. However the balance of opinion appeared to favour the underlying idea that revealing other people's energy usage data would lead to competition and peer

influence to reduce consumption. We are currently developing the experimental design of a real system which uploads energy consumption data direct to a remote server so that Facebook (and other applications) can potentially make use of such data.

Further embryonic work has involved the use of a recently developed website, myenergyusage.com, which employs the use of a desktop application called Powometer [26]. This application has the capability to send monitoring data online from Wattson allowing members of the website to view their individual, as well as group collective, energy usage. Participant energy usage is graphically displayed using various gauges and graphs representing both real time and historic energy data. Work in progress involves the linking of the energy monitoring data to the Facebook application using open standards. It is expected that using open standard approach will pave the way for many different types of application development such as desktop widgets, mobile phone applications and RSS feeds. We believe that opening the data to as many platforms as possible is a primary step in creating more awareness of domestic energy consumption.

REFERENCES

[1] Iyer, M., Kempton, W., and Payne, C., (2006) Comparison groups on bills: Automated, personalized energy information. *Energy and Buildings*. 38(8): 988-996.

[2] Owen, G., and Ward, J., (2006) *Smart meters: commercial, policy and regulatory drivers*. Sustainability First, London

[3] 2 Save Energy PLC, (2009) *Save Electricity - OWL Electricity Monitor saves you money*, available from **http://www.theowl.com/**

[4] Eco-Eye, (2009) ***Eco-Eye** Real Time Electricity Monitors*, available from **http://www.eco-eye.com/**

[5] DIY Kyoto, (2009) DIY *KYOTO – Wattson*, available from **http://www.diykyoto.com/**

[6] Pierce, J., and Roedl, D., (2008) Changing Energy Use Through Design. *Interactions Magazine*, XV.4 -,. Pp 6-12.

[7] Arrington, M., (2009) "*Facebook Now Nearly Twice The Size Of MySpace Worldwide*" TechCrunch . vailable from **http://www.techcrunch.com/2009/01/22/facebook-now-nearly-twice-the-size-of-myspace-worldwide/.**

[8] Prophetic Media, (2009) *AppData.com - Facebook Application Metrics*, available from **http://www.appdata.com/**

[9] Mankoff, J., Matthews, D., Fussell, S.R., and Johnson, M., (2007) Leveraging Social Networks to Motivate Individuals to Reduce their Ecological Footprints, *Proceedings of HICSS*, pp. 87.

[10] Nazir, A., Raza, S., and Chuah, C., (2008) Unveiling Facebook: a measurement study of social network based applications. In *Proceedings of the 8th ACM SIGCOMM Conference on internet Measurement* (Vouliagmeni, Greece, October 20 - 22, 2008). IMC '08. ACM, New York, NY, pp 43-56

[11]Triandis, C.T., (1971) *Attitude and Attitude Change*: London, Wiley

[12] M. Weiksner, B. J. Fogg, X. (2008) Liu: Six Patterns for Persuasion in Online Social Networks. *Proc of PERSUASIVE*: pp151-163

[13] Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press

[14] Moscovici, S. and Zavalloni, M. (1969). The Group as a Polarizer of Attitudes. *Journal of Personality and Social Psychology*, *12*, pp 125-135.

[15] Stoner, J. A. F. (1961). *A Comparison of Individual and Group Decisions Involving Risk.* Unpublished Master's Thesis, Massachusetts Institute of Technology.

[16]Sia, C. L., Tan, B. C. Y., and Wei, K. K., (2002). Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity. *Information Systems Research*, 13, 1, 70-90

[17] Facebook, (2009) *Feed a Child with just a Click!:* available from , **http://www.facebook.com/group.php?gid=6324544002**

[18] Kitayama, S., Burnstein, E., (1994) Social Influences, Persuasion, and Group Decision Making. In Brock, T.C., ed.: *Persuasion: Psychological Insights and Perspectives.* Allyn and Beacon 175 – 194

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

[20] NHS, (2009) *Stop Smoking and Start Living*, available from **http://smokefree.nhs.uk/quit-tools/make-a-promise/**

[21] Quit Smoking Journals, (2009) *Quit Smoking Journals – Free Quit Smoking Group*, available from **http://www.quitsmokingjournals.com/**

[22]Dillahunt, T., Becker, G., Mankoff, J., and Kraut, R., (2008) Motivating environmentally sustainable behavior changes with a virtual polar bear. *Proc of PERSUASIVE*, pp.58–62.

[23] Cairns, P., and Cox, A.L, (2008) *Research Methods for Human-Computer Interaction*. Cambridge, Cambridge Press

[24]Strauss, A., (1987) *Qualitative analysis for social scientists*: Cambridge, Cambridge University Press

[25]Bang, M., Torstensson, C., and Katzeff, C., (2006) The PowerHouse: A Persuasive Computer Game Designed to Raise Awareness of Domestic Energy Consumption *Lecture Notes in Computer Science*, Vol. 396, pp. 123-132.

[26] Proctor C., (2009) *My Energy Usage – Electricity Monitoring Software*, available from **http://www.myenergyusage.org/**

Using On-board Driver Feedback Systems to Encourage Safe, Ecological and Efficient Driving: The Foot-LITE Project

R G Fairchild[1], J F Brake[1], N Thorpe[1], S A Birrell[2], M S Young[2], T Felstead[3] and M Fowkes[4]

Abstract. In response to increasing political and individual awareness of the need to address the social and environmental costs of unsafe, inefficient and highly polluting driving styles, the Foot-LITE research project seeks to deliver innovative driver/vehicle interface systems and services to encourage and hopefully persuade sustained changes to driving styles and wider travel behaviour. Stakeholders' requirements help to define the functionalities of the system being developed in the context of a rapidly evolving market with many products potentially competing for Original Equipment Manufacturer (OEM) application or retrofitting to vehicles.

1 INTRODUCTION

The Foot-LITE project addresses two of the Future Intelligent Transport Systems' (FITS) Innovation Platform priorities in a joint call from the UK Department for Transport (DfT), the Technology Strategy Board (TSB) and the Engineering and Physical Sciences Research Council (EPSRC) in October 2006: namely, improving road safety and better travel and traveller information. The accepted Foot-LITE proposal included a case for an eco-driving and safety advisory system. It intends to deliver innovative driver/vehicle interface systems and back office services to encourage sustained changes to driving styles and behaviours which are safer, reduce congestion, enhance sustainability, help reduce traffic pollution emissions and reduce other social and environmental impacts. These objectives are encompassed by three main characteristics of altered driving behaviour that the Foot-LITE system intends to deliver: eco-friendly (green), safe and efficient driving. The Foot-LITE project began in the summer of 2007 and is currently due to be completed by mid-2010.

There are clear interdependencies between the three driving styles noted above; for example, eco-driving can be a by-product of safe driving as it leads to a better anticipation of events, whilst driving and network efficiencies are derived from safer and greener driving styles. However, the different driving styles may not always be complementary. Finally, analysis of the driving task suggests that the Foot-LITE design must be acceptable, useable, non-distracting and be a valued interface that supports the driver.

[1] TORG, School of Civil Engineering & Geosciences, Newcastle University, NE1 7RU, UK. Email {**r.g.fairchild**, j.f.brake, neil.thorpe} @ncl.ac.uk
[2] Ergonomics Research Group, Brunel University, Uxbridge, UB8 3PH, UK. Email {stewart.birrell, m.young} @brunel.ac.uk
[3] TRG, School of Civil Engineering & the Environment, University of Southampton, SO17 1BJ, UK. Email t.j.felstead@soton.ac.uk
[4] MIRA, Nuneaton, CV10 0TU, UK. Email mark.fowkes@mira.co.uk

2 MARKET ANALYSIS

There has been increasing awareness of the impact of vehicle emissions on the environment combined with a period of rapidly increasing fuel prices in the latter part of 2008. This has encouraged several vehicle manufacturers and independent organisations to produce devices and services with similar motivations to Foot-LITE.

These systems and services do not fulfil all the characteristics of driving that Foot-LITE addresses. An objective of the Foot-LITE system is to provide both online and offline driver feedback for safety and eco-driving, which is not offered by the systems reviewed. Fiat's eco:Drive system [1] requires the user to download vehicle data using a USB drive; this data set is then uploaded to the user's home computer where it is analysed using a Fiat webservice, providing only offline feedback. The eco:Drive product provides information such as journey cost and CO_2 emissions, alongside tips for better driving. Green Road [2] is a system that collects in-vehicle data and analyses it to provide both on- and off-board feedback. The on-board feedback is presented by a simple dashboard display that shows a green, amber or red light depending on the driving style, with red indicating riskier and/or more polluting behaviour. More detailed analysis is presented by text message to the user's mobile phone and on the Green Road website. The drawback to the Green Road method of on-board information presentation is that it does not provide information on the specific behaviour causing the poor driving performance at that instance and therefore the driver does not instantly learn the corrective behaviours.

Other products such as the Driver Fatigue Monitor (DFM) produced by SafeDrive Europe [3] only alert the driver to specific safety issues, in this case the warning signs of a driver falling asleep. A system soon to be available on new Vauxhall Insignia models provides road sign recognition along with lane departure warnings [4]. Specific issues are addressed with specialist products such as Audi's 'Travolution' project [5]. This system interrogates traffic signals as the vehicle approaches and if the traffic signal is red, the system can indicate the ideal speed for the vehicle to travel from that point in order to reach the lights as they turn green. This reduces the need for braking and accelerating, thereby creating smoother journeys and reduced emissions.

There are also services available for training drivers in safe and economical driving. These are mostly aimed at fleet users

and professional drivers of the largest and most polluting vehicles such as city buses and heavy goods vehicles. Several companies large and small offer these services including the AA, the Institute of Advanced Motorists and the Royal Society for the Prevention of Accidents (RoSPA). Whilst these training courses garner successful results in the short-term by correcting poor driving styles, they do not provide continuous driving behaviour feedback that an in-vehicle device such as Foot-LITE presents. Furthermore, whilst being aimed at fleet users, there is no mechanism to feedback to the fleet manager after the initial course monitoring of continuous driving behaviour, essentially allowing drivers to revert back to old habits should they wish.

3 CONCEPT DEVELOPMENT

Cognitive Work Analysis (CWA) [6] was carried out in order to define a framework for identifying the parameters of the Foot-LITE system [7]. This leads to the identification and definition of variables known as Physical Objects that can be used to describe the Foot-LITE system; they are labelled as Physical Objects to encompass the measurement of physical metrics of the vehicle (such as speed) and describe physical attributes of the device (such as display screen size). The Physical Objects were assigned to one or more situations in a Contextual Activity Template (CAT): pre-driving, low mental workload driving, high mental workload driving, immediate post-driving and an evaluation period after the journey. In addition, Use Case Analysis was conducted as a parallel activity, which identified three further stages: organisational and legal, establishing the system and system decommissioning.

Three main stakeholder groups were identified. The first group, the consortium stakeholders, comprised the sponsors, product manufacturers and consortium partners. Second, interested organisations included policy makers, vehicle manufacturers and motoring organisations. Finally, end user groups were categorised according to private organisations or local/statutory authorities with a bulk fleet management function, driving schools and individuals.

Stakeholder requirements, conducted by Newcastle University, identified the attitudes of the stakeholder groups using a mixed methods approach. Research project members provided knowledge input through the CWA and a series of structured sessions, whilst interested organisations contributed through the CWA and focus groups. Potential end users were consulted via focus groups and questionnaires. The aims of the focus groups were twofold: to confirm or reject the knowledge characteristics generated by the CWA and to identify additional knowledge characteristics that had been rejected or not identified by the CWA. Information from potential end users was required about the knowledge requirements, rather than the device requirements, which enable the technical realisation of the knowledge. Focus groups sampled across the potential market. The first step was to identify the key characteristics of individuals with a pre focus group questionnaire.

At the focus groups background statements were used to provide context by describing the objectives of the system, examples of safe, green and efficient driving based on an advanced driving guide and potential characteristics of the system. The format of the focus groups was structured using the CAT. For each stage, a general open question was asked about the types of information attendees would like to help them with their driving, followed by closed questions related to the Physical Objects. This procedure was followed for safe, green and then efficient driving before considering the next stage. Low and high mental workload states were dealt with together, as attendees were able to understand better the different implications of these two stages. Finally, participants completed a short questionnaire in order to assess the system's market potential, when they would use it and the physical appearance of the in-vehicle device.

The outcomes of the literature review, CWA and focus groups were intended to inform the subsequent stages of the project. First, the need for Foot-LITE and potential applications of the system were identified. This began with a market review of competing systems (as detailed in Section 2) that identified how Foot-LITE could gain a commercial advantage. The potential short/medium term applications of the system were identified, starting with a review of key reports such as the Stern Review, the Eddington Report and the King Reports to identify policy areas that Foot-LITE could address [8], [9], [10], [11]. The environmental context of Foot-LITE was then assessed in relation to climate change, vehicle emissions, local pollutants and noise. The individual context considered how to appeal to end-users by demonstrating economic, ethical and other personal benefits. The organisational context built upon the individual context and demonstrated why targeting Foot-LITE at fleet users may encourage a more rapid take-up of the system, since organisations must address political, legal and financial challenges by delivering improved fleet management through back office functions. Societal benefits considered broader environmental issues relating to driver behaviour and training.

End User Benchmarks were defined with the premise that the Foot-LITE driving style will benefit network performance. Engine speed, gear choice and throttle position were the most important parameters to be addressed. The Institute of Advanced Motorists (IAM) have produced a driving guide [12] which is proposed as the benchmark for Foot-LITE, and hence the Foot-LITE drivers, to achieve.

4 Foot-LITE FEEDBACK

The Foot-LITE system will deliver feedback to drivers and potentially vehicle owners (if not the driver) in order to promote the take-up and retention of appropriate eco-friendly (green), safe and efficient driver behaviour [7].

Various driver behaviours and attitudes, including attitudes towards car maintenance, are encompassed in each area. Eco-friendly driving can be seen as a method of reducing fuel consumption and therefore (potentially) emissions. Several studies have investigated the fuel consumption of drivers before and after training. Depending on the type of vehicle used and the initial skill of the driver, figures from a review of studies show a reduction in fuel consumption typically in the order of 6 to 15% [14], [15], [16]. Eco-driving encompasses several characteristics, some of which can be altered by the use of a system like Foot-LITE and are listed below:

- driving style (the primary focus of Foot-LITE);
- vehicle maintenance;
- vehicle loading;
- vehicle drag; and
- the use of vehicle accessories.

Safe driving encompasses the ability of the driver to control the vehicle in all conditions, adherence to the laws of the road and knowledge of the adaptation of the normal driving style in adverse environmental conditions. Several studies have shown that driving speed is the main factor in the risk and severity of a crash [17], [18], [19]. Safe driving and eco-driving can, in some circumstances, be complementary. For example, by anticipating traffic signals and slowing accordingly in order to avoid stopping the vehicle, the driver is also approaching queuing traffic more cautiously and, should the need for rapid braking occur, is already travelling at a slower speed. However, in some circumstances eco- and safe driving may not be in harmony; for example, trying to maintain a constant speed by avoiding braking may compromise vehicle headway; also, travelling in the highest possible gear may adversely affect vehicle control. Thus, in some situations, priority may have to be given to safe rather than eco-friendly driving behaviours [13].

Within the context of Foot-LITE, the aspect of efficiency refers to efficient road transport, this being the generalised cost for both the Foot-LITE user (individual) and other road users (network impacts). This is achieved in the Foot-LITE system by informing the driver before and during a trip about the state of the road network. This includes suggestions to undertake the journey at another time or using a different mode (by providing appropriate pre-trip information) or by providing alternative routing that avoids traffic congestion or even cancelling the journey altogether. If a driver is already committed to a journey by car, during-trip information would use dynamic traffic information to present alternative choices such as a revised route and its associated journey time.

5 HUMAN MACHINE INTERFACE

Modern vehicles contain an increasing amount of instrumentation as a combined consequence of factors including the motivations of vehicle manufacturers, advances in technology, government legislation and consumer demand. Whilst in-vehicle Human Machine Interfaces (HMI) have existed for almost as long as motorised vehicles themselves, in the last 15 years these interfaces have evolved beyond simple dashboard instrumentation and in-car audio equipment. The rapid development of Advanced Driver Assistance Systems (ADAS) has resulted in an increased need for these systems to interact with and inform the driver. However, the added information available to the driver raises significant ergonomic concerns for driver mental workload and driving task performance.

Driver mental workload can be affected by a number of factors, which are a combination of being external (e.g. traffic, road conditions) and internal (e.g. the driver's age and experience) to the individual [20], [21]. In addition, different elements of the driving task (such as navigation and vehicle control) can impose varying levels of mental workload. For instance, steering appears to be a significant source of workload in vehicle control [22], whilst tuning a car radio is one of the most demanding in-car tasks [23].

6 THE Foot-LITE SYSTEM

In order to define achievable final user requirements for the Foot-LITE system, the outputs from the end user requirements were considered together with the other stakeholder opinions. These were classified as general functionality, advice functionality and metrics (driver and specific).

General functionality ensures that the system meets non-advice criteria, e.g. the requirement for security means the system needs to identify drivers and ensure that data uploaded to a back-office system for analysis is kept confidential and secure. Configurability is very important, for example, the selection of driver preferences and focusing on weak points of driving style. Some general functions were not derived directly from the end user requirements. For example, the system has to operate during all times of the day and adjust advice according to the external environment. Furthermore, all data has to be stored in a driver specific manner and advice to the driver must be timely, enabling the driver to take action that will improve driving performance.

Advice functionality describes what advice is presented to the driver and how this advice is disseminated. Thus, the system can utilise visual, audible and tactile means of providing advice and training to the driver, whilst not knowingly giving advice which is unsafe or presents problems for other road users. Advice may include the appropriate lane position with respect to other road users and road features and identification of inappropriate lateral movement.

On- and off-line driver metrics are predefined levels against which the driver's behaviour is measured. They are required in order to fulfil the general and specific functionalities. The system must be able to aggregate data over time in order to monitor driver performance and store data locally or transmit data to an off-vehicle system for further analysis and storage. It also needs to feed selected metrics back to the driver in real time (on-line) as an immediate measure of performance and off-line in order to enable the user to examine trends in their behaviour over time.

7 SUMMARY

Foot-LITE, as an encouraging and possibly persuasive digital technology, will monitor drivers' behaviour, vehicle metrics and road network conditions. These data sources will be analysed via an on-board device, providing information for appropriate advice, tips and useful reminders to be presented to the driver with consideration to the journey stage and the driver's mental workload state.

The Foot-LITE project has reached the stage of human machine interface (HMI) design with two system designs being simulated in software, which will be tested using volunteers on a full scale vehicle simulator at Brunel University. The approved design, after revisions based on user feedback, will be constructed in hardware integrated into a user device with vehicle data inputs, touch sensitive liquid crystal display (LCD) and audible alert capability ready to be manufactured for further trials.

Following the trials of the on-board Foot-LITE device on the Brunel simulator, it will be applied to a highly instrumented vehicle at Southampton University before moving to a small fleet of three vehicles and finally onto a fleet of up to 30 vehicles.

Evaluation of the results of the trials will assess the suitability of the input data gathering, data analysis and comparative driver behaviour models and advice/commands output methods. As

well as evaluating the physical and analytical characteristics of the device, the trials will also assess the feasibility of such a system to change driver behaviour not only in terms of their technical competence for achieving and maintaining safe, ecological and efficient driving, but also personal acceptance and wider benefits of the technology.

REFERENCES

[1] Fiat (2009) *Fiat eco:Drive,* Available from: <http://www.fiat.com/ecoDrive/>

[2] Green Road Technologies (2009) *Green Road Technologies – Fleet Safety – Driver Safety.* Available from: <http://www.greenroad.com/>

[3] SafeDrive Europe Ltd. (2008) *SafeDrive Europe Ltd.* Available from: <http://www.safedriveeurope.com/products.html>

[4] Eden Vauxhall (2008) *Insignia First – New Vauxhall Technology.* Available from: <http://www.vauxhall-insignia.com/News/vauxhall-insignia-camera-system.aspx>

[5] Audi (2008) *Audi Backed 'Travolution' Project gets the Green Light.* Available from: <http://www.newspress.co.uk/DAILY_LINKS/arc_sep_2008/190908aud.htm>

[6] Vicente, K. (1999) *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*, Mahwah, NJ: Lawrence Erlbaum Associates.

[7] Birrell, S.A., Young, M.S., Stanton, N.A. and Jenkins, D.P. (2008). *Improving driver behaviour by design: A Cognitive Work Analysis methodology.* Applied Human Factors and Ergonomics 2nd International Conference, 14-17 July 2008, Las Vegas, USA.

[8] Stern, N. (2006) *Stern Review on the Economics of Climate Change.* London, HM Treasury. 30th October. Available from: <http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/sternreview_index.cfm>

[9] Eddington, R. (2006) *The Eddington Transport Study. Department for Transport.* 1st December. Available from: <http://www.dft.gov.uk/about/strategy/transportstrategy/eddingtonstudy/>

[10] King, J. (2007) *The King Review of Low-carbon Cars - Part 1: The Potential for CO2 Reduction*. London, HM Treasury. 9th October. Available from: <http://www.hm treasury.gov.uk/pbr_csr/reviews/pbr_csr07_king_index.cfm>

[11] King, J. (2008) *The King Review of Low-Carbon Cars - Part 2: Recommendations for Action*. London, HM Treasury. 12th March. Available from: <http://www.hm-treasury.gov.uk/budget/budget_08/reviews/bud_bud08_king.cfm>

[12] Institute of Advanced Motorists (2007) *Advanced Driving – The Essential Guide.* MBI Publishing Company.

[13] Young, M.S., Birrell, S.A., Stanton, N.A., Thorpe, N. and Fowkes, M. (2008) *Safe and fuel efficient driving: Defining the benchmarks* In P. D. Bust (Ed.), Contemporary Ergonomics 2008 (pp. 749-754). London: Taylor & Francis.

[14] De Vlieger, I. (1997) *On-board emission and fuel consumption measurement campaign on petrol-driven passenger cars*. Atmospheric Environment, 31(22), 3753-3761.

[15] Johansson, H., Farnlund, J. and Engstrom, C. (1999) *Impact of EcoDriving on emissions and fuel consumption, a pre-study*. Borlange, Sweden: Swedish National Road Administration (pp. 1-41).

[16] IEA/OCEAD IEA (2005) *Making Cars More Fuel Efficient*. IEA/OCEAD. Accessed: 17/01/08. Available from: <http://iea.org/textbase/nppdf/free/2005/fuel_efficient.pdf>

[17] Haworth, N. and Symmons, M. (2001) *The relationship between fuel economy and safety outcomes*. Report No. 188 (pp. 1-57). Australia: Monash University, Victoria.

[18] Taylor, M., Baruya, A. et al (2002) *The relationship between speed and accidents on rural single-carriageway roads.* Wokingham: TRL (pp. 1-32).

[19] Aarts, L. and van Schagen, I. (2006) *Driving speed and the risk of road crashes: A review. Accident Analysis & Prevention*, 38, 215-224.

[20] Schlegel, R. E. (1993) *Driver mental workload.* In B. Peacock and W. Karwowski (Eds.), Automotive ergonomics. (pp. 359-382). London: Taylor & Francis.

[21] Verwey, W. B. (1993) *How can we prevent overload of the driver?* In A. M. Parkes and S. Franzen (Eds.), *Driving future vehicles*. (pp. 235-244). London: Taylor & Francis.

[22] Young, M. S. and Stanton, N. A. (1997) *Automotive automation: Investigating the impact on drivers' mental workload.* International Journal of Cognitive Ergonomics, 1, 325-336.

[23] Dingus, T. A., Antin, J. F., Hulse, M. C. and Wierwille, W. W. (1989) *Attentional demand requirements of an automobile moving-map navigation system*. Transportation research. Part A: general.

The partners in the Foot-LITE project are: MIRA Ltd, Auto-txt, Hampshire County Council, TRW Conekt, Institute of Advanced Motorists, Ricardo UK Ltd, Transport for London Ltd, Zettlex Printed Technologies Ltd, University of Southampton, Brunel University and Newcastle University.

The authors would like to acknowledge the support of the sponsors of this research namely the Department for Transport, the Technology Strategy Board and the Engineering and Physical Sciences Research Council in the UK.

Introduction to the LifeGuide: software facilitating the development of interactive behaviour change internet interventions

Lucy Yardley[1], Adrian Osmond[2], Jonathon Hare[2], Gary Wills[2], Mark Weal[2], Dave de Roure[2] and Susan Michie[3]

Abstract. We are developing a set of software resources named 'the LifeGuide' that will enable researchers to collaboratively create, evaluate and modify two central dimensions of behavioural interventions: a) providing tailored advice; b) supporting sustained behaviour.

1 INTRODUCTION: CONTEXT FOR THE LIFEGUIDE

Behavioural interventions - packages of advice and support for behaviour change - are arguably the most important methodology and technology employed by social scientists for understanding and changing behaviour. The applied value of behavioural interventions is that they can promote and support behaviour that benefits the individual and society. Their scientific value is that behavioural interventions provide the strongest test of theories that seek to identify the causes of behaviour [1]. Whereas observational studies can only note that certain factors (e.g. beliefs, attitudes, skills) are associated with desired behaviour changes, behavioural interventions can experimentally test whether changing these factors causes a change in behaviour. Behavioural interventions can be used for a wide range of very different behaviours; for example, to change risky or antisocial behaviour, improve productivity and reduce accidents in the workplace, enhance learning activities, or promote environmentally important behaviour change, such as reducing energy use.

Despite their crucial importance, to date progress has been disappointingly slow in developing effective behavioural interventions [1, 2]. There has also been a lack of cumulative theoretical development about what components of interventions work, in which combinations and for whom [3, 4]. These problems are partly due to the high costs of large trials of interventions with long-term follow-up, which has resulted in a dearth of studies that are adequately powered to address these key theoretical questions.

Internet-based behavioural interventions are set to play a crucial role in the delivery and evaluation of behavioural interventions in the near future, for reasons detailed below. The techniques of e-Science provide a tremendous opportunity to support both the delivery of these interventions and the associated research, while internet-based behavioural interventions demand innovation in e-Science in order to develop the operational and research capability for real-time interactive engagement with a large number of users.

2 THE ADVANTAGES OF INTERNET-BASED BEHAVIOURAL INTERVENTIONS

Behavioural interventions have traditionally been delivered principally face-to-face, and this continues to be the overwhelmingly dominant method of delivery. A major problem with this mode of delivery is that it is extremely resource intensive, severely limiting the scope for cost-effective interventions; clearly, it is not feasible to provide every individual with 24 hour access to personal advice and support for managing all aspects of their lives.

In contrast, internet-based behavioural interventions can be made available to most of the population for little more than the cost of development [5, 6]. Whereas the quantity and timing of information, advice and support that can be delivered face-to-face is very restricted, internet-based behavioural interventions can be accessible at all times and provide extensive and intensive advice and support. Currently delivered principally over the web, internet-based behavioural interventions will increasingly be flexibly accessible through mobile phones, interactive digital TV etc.

There are two key dimensions of behavioural interventions. The 'motivational' dimension involves providing relevant information, advice, education and decisional aids in order to promote knowledge, beliefs, attitudes and intentions consistent with the desired behaviour. However, a 'volitional' dimension is often also vital in order to help people translate good intentions into behaviour [7-9]. Consequently, effective behavioural interventions also provide a variety of techniques to support and sustain behaviour change, such as aids to goal-setting, planning and self-monitoring, skill and confidence-building, cues and reminders, and systems of incentive and social support [10-13]. Since face-to-face delivery is not cost-effective for many behavioural interventions, printed materials (e.g. booklets and manuals) are often used to disseminate motivational behavioural interventions more widely. However, the information and advice provided in printed materials cannot easily be 'tailored' or customised to the particular situation of the individual, and may therefore be dismissed as irrelevant [14]. Moreover, even when booklets are effective in changing knowledge and beliefs, they cannot provide the volitional dimension of ongoing interactive support for behaviour change, such as prompts, feedback and encouragement.

[1] School of Psychology, Univ. of Southampton, SO17 IBJ, UK. Email: L.Yardley@soton.ac.uk

[2] Electronics and Computer Science, Univ of Southampton, SO17 1BJ UK. Email: {aco|jsh2|gbw|mjw|dder}@ecs.soton.ac.uk

[3] Department of Psychology, Univ College London, WC1E 7HB, UK. Email: S.Michie@ucl.ac.uk

Interactive internet-based behavioural interventions can provide information and advice specifically 'tailored' to address the particular situation, concerns, beliefs and preferences of the individual, and may therefore be more persuasive than generic printed information [14, 15]. Interactive internet-based behavioural interventions can also provide a rich, stimulating, engaging and actively supportive environment, with audiovisual illustrations, reminders, personalised feedback regarding progress and concerns, and opportunities for peer-to-peer support and comparison [16, 17]. Supporting long-term maintenance of desirable behaviour changes is a major problem that has not yet been solved [18]. Internet-based behavioural interventions may for the first time offer a cost-effective means of providing long-term support.

Currently, the potential of internet interventions is restricted by the lack of software that will allow researchers to easily create all aspects of an interactive intervention, without the need to programme the software infrastructure for each intervention individually. This limits the number of interventions that can be developed and evaluated, thereby limiting the accumulation of knowledge about intervention effectiveness, and the relative effectiveness of intervention components and causal mechanisms. The initial development costs are typically greater for internet interventions than for traditionally delivered interventions, and once programmed they cannot easily be modified, acting as a barrier to innovation and enhancement of interventions. Lack of access to resources for programming interventions also restricts the numbers of researchers that can engage in developing and testing internet interventions, and in particular makes it more difficult for postgraduate students and junior researchers to engage in this type of research. Commercial software packages have been developed recently that allow professional users some scope to enter the content for the particular intervention that they wish to create, but these restrict the researcher to those components pre-selected by the developers, and do not offer the research community the crucial advantages of free access and the ability to innovate methods and integrate findings.

Since the essential components, functions and underlying infrastructure required for internet interventions are common to a vast range of applications, it makes sense to develop an open-access set of shared software resources that researchers can use to easily create and modify different interventions themselves – the LifeGuide.

3 FUTURE OF THE LIFEGUIDE

The LifeGuide is being designed to provide the research community with facilities to collaboratively devise complex interventions, with immediate access to components that have been validated in previous LifeGuide projects, that can then be utilised (modified as necessary) in new applications. The research community will then be able to work together to rapidly recruit participants from geographically dispersed locations, and integrate the data to form very large data-sets that can be used to carry out more powerful analyses than have hitherto been possible, such as mediator and moderator analyses of intervention effects.

As internet interventions become very widely used in many spheres of life, this could in the future provide the basis for national and international 'population laboratories' for the continuous further refinement of interventions. Ultimately, a semantically enriched and adaptive LifeGuide system should be able to continuously and semi-autonomously model and refine interventions, based on the preferences and outcomes of lay users. We also plan to ensure that LifeGuide interventions can interface with existing medical systems (e.g. patient records), and with remote monitoring devices (e.g. monitoring blood glucose levels, physical activity, heart functioning etc.), and can be delivered via a range of digital media (e.g. mobile phone, television).

4 THE LIFEGUIDE: WORK IN PROGRESS

We are using a co-design approach to ensure that the LifeGuide offers the flexibility needed to deliver a very wide range of interventions with different formats and ingredients, and is sufficiently user-friendly that novice researchers can readily use it to develop interventions, with the aid of the basic online training package, manual and help we are developing. An international network of researchers has been recruited through our workshops and demonstrations; these researchers are collaborating with us in evaluating and developing the LifeGuide by applying it to a range of very different health problems. The interventions include: an 'internet dr' intervention to provide people suffering from common conditions (e.g. colds and 'flu, irritable bowel syndrome) with tailored advice that enables them to cope with their symptoms; an e-learning website for health trainers to teach them how to help clients adopt healthy behaviour; and an intervention to promote and support hygienic behaviour to reduce the spread of infection, especially during pandemic flu. We are currently carrying out qualitative and quantitative pilots of these interventions, prior to conducting substantive tests of them.

The LifeGuide software has two main parts, the authoring tool and the LifeGuide server software.

The authoring tool has been developed using Eclipse RCP (Rich Client Platform) technology and allows intervention authors to construct individual pages, in terms of content and presentation, and link them together using the intervention logic. Logic can be based on current user selections, or data from past sessions. The intervention pages created by the authoring tool can be saved, loaded and edited locally as the author constructs their intervention. The intervention files are held in an xml format promoting extensibility and interoperability. Figure 1 illustrates the user-interface of the authoring tool in page-editing mode, and shows the level of graphic design achievable with the tool.

Once the intervention has been constructed, the XML files can be exported from the authoring tool into a single intervention file that can then be uploaded to the LifeGuide server. The LifeGuide server allows researchers to run trials of their authored interventions and collect the trial data in a secure database allowing authorized access at various levels of anonymisation. The server uses technology developed from the ASDEL project (Assessment Delivery Engine for QTIv2)[4] [19] and XSLT style templates to render the intervention as a series of XHTML pages viewed by the participants of the intervention using standard browser technology. By using Web standards,

[4] http://www.asdel.ecs.soton.ac.uk/

efforts have been made to ensure browser compatibility and the use of XSLT will allow the continued development of delivery to a wider range of mobile and embedded browser platforms.

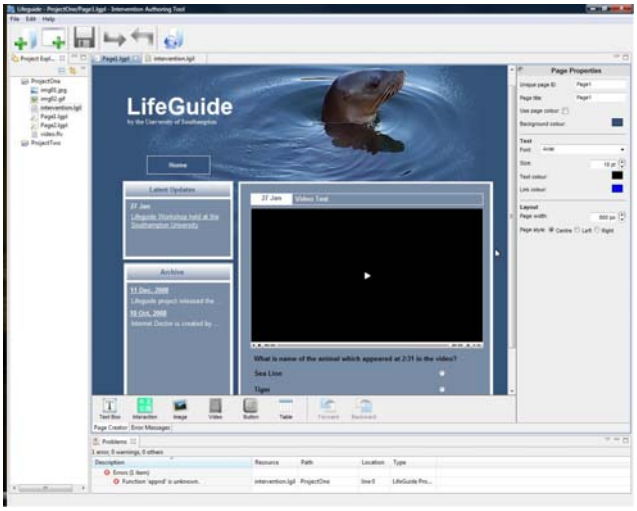

Figure 1: The LifeGuide authoring tool

Using LifeGuide, researchers are able to select and change the question types and response options used to assess lay user status, and the content and format of the tailored advice delivered (which is matched to user responses). The flexible interface gives researchers the ability to select and change multimedia resources linked to the site (e.g. audiovisual resources illustrating symptoms, self-care techniques etc.), adjust the look and feel of the system (e.g. colours, font, skins); and design their own user interface, using templates. LifeGuide will also facilitate collaboration in intervention development (e.g. through discussion boards and links to video-conferencing facilities). LifeGuide collects output data on participant use and outcomes, stores it securely, and provide facilities for collating and outputting anonymised data. Figure 2 illustrates the manager software which gives user details such as how many people used a particular intervention in a given time, the actual order of the pages viewed and where the users were geographically located.

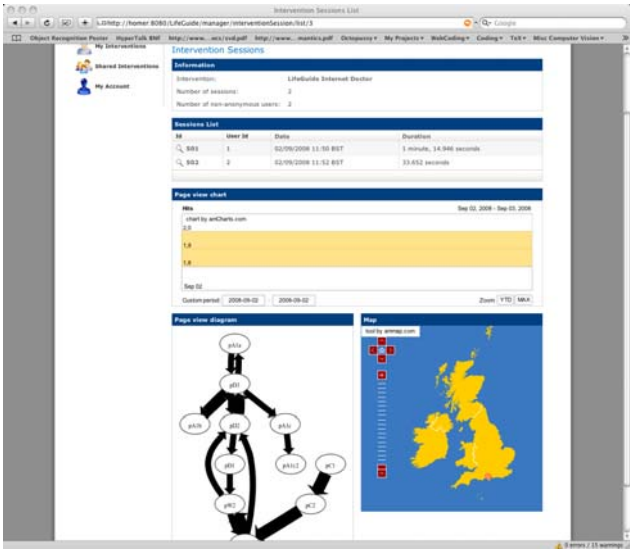

Figure 2. Features of the intervention manager software

Trial data can be output in SPSS and Excel (csv) formats and a wider variety of formats can be supported as and when the need arises.

Coding standards have been adopted to ensure readability and testability. Full account has been taken of issues relating to accessibility of Web-based systems and software to people with disabilities and the outputs of this project will conform to published standards and guidelines (e.g., the WC3 Web Accessibility Initiative level Double-A).We are endeavouring to future-proof the LifeGuide software by using open source software and adopting open standards that are technology independent, facilitating easy updating. We will also ensure that data collected is not 'locked-in' to the system or approaches used, enabling other systems to reuse the data.

More details of LifeGuide and how to access it can be found at: www.lifeguideonline.org

ACKNOWLEDGEMENTS

LifeGuide is funded by the UK Economic and Social Research Council (ESRC) as part of the National Centre for e-Social Science (NCeSS).

REFERENCES

[1] S. Michie, A. J. Rothman, and P. Sheeran. Advancing the science of behaviour change. *Psychology and Health*, 22:249-253 (2007).

[2] A. Warsi, P. Wang, M. P. LaValley, J. Avorn, and D. H. Solomon. Self-management education programs in chronic disease: a systematic review and methodological critique of the literature. *Archives of Internal Medicine*, 164:1641-1649 (2004).

[3] S. Michie and C. Abraham. Interventions to change health behaviours: evidence-based or evidence-inspired? *Psychology and Health*, 19:29-49 (2004).

[4] P. M. Nicassio, B. E. Meyerowitz, and R. D. Kerns. The future of health psychology interventions. *Health Psychology*, 23:132-137 (2004).

[5] P. A. Jennett, H. L. Affleck, D. Hailey, A. Ohinmaa, C. Anderson, R. Thomas, B. Young, D. Lorenzetti, and R. E. Scott. The socio-economic impact of telehealth: a systematic review. *Journal of Telemedicine & Telecare*, 9:311-320 (2003).

[6] F. Griffiths, A. Lindenmeyer, J. Powell, P. Lowe, and M. Thorogood. Why are health care interventions delivered over the internet? A systematic review of the published literature. *Journal of Medical Internet Research*, 8:e10 (2006).

[7] C. J. Armitage and M. Conner. Social cognition models and health behaviour: a structured review. *Psychology and Health*, 15:173-189 (2000).

[8] F. F. Sniehotta, U. Scholz, and R. Schwarzer, Bridging the intention-behaviour gap: planning, self-efficacy and action control in the adoption and maintenance of physical exercise. *Psychology and Health*, 20:143-160 (2005).

[9] P. Sheeran. Intention-behaviour relations: a conceptual and empirical review. In: *European Review of Social*

Psychology, 12. W. Stroebe and M. Hewstone. Eds. Chichester: Wiley, pp. 1-36 (2002).

[10] P. M. Gollwitzer and P. Sheeran. Implementation intentions and goal achievement: a meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38:69-119 (2006).

[11] I. C. A. Hobbis and S. Sutton. Are techniques used in cognitive behaviour therapy applicable to behaviour change interventions based on the theory of planned behaviour? *Journal of Health Psychology*, 10:7-18 (2005).

[12] P. Karoly. Mechanisms of self-regulation: A systems view. *Annual Review of Psychology*, 44:23-52 (1993).

[13] A. Prestwich, M. Conner, R. Lawton, W. Bailey, J. Litman, and V. Molyneaux. Individual and collaborative implementation intentions and the promotion of breast self-examination. *Psychology & Health*, 20:743-760 (2005).

[14] M. Kreuter, D. Farrell, L. Olevitch, and L. Brennan. *Tailoring health messages: customizing communication with computer technology*. London: Erlbaum, (2000).

[15] H. De Vries and J. Brug. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. *Patient Education and Counseling*, 36:99-105 (1999).

[16] B. G. Danaher, S. M. Boles, L. Akers, J. S. Gordon, and H. H. Severson. Defining participant exposure measures in Web-based health behavior change programs. *Journal of Medical Internet Research*, 8:e15 (2006).

[17] C. Kerr, E. Murray, F. Stevenson, C. Gore, and I. Nazareth. Internet interventions for long-term conditions: patient and caregiver quality criteria. *Journal of Medical Internet Research*, 8:e13 (2006).

[18] A. J. Rothman. Toward a theory-based analysis of behavioral maintenance. *Health Psychology*, 19:64-69 (2000).

[19] G. Wills, H. Davis, L. Gilbert, J. Hare, Y. Howard, S. Jeyes, D. Millard, and R. Sherratt. Delivery of QTIv2 Question Types. In: Procs. Int'l Conference, Loughborough, UK (2007).

Website credibility and intervention effectiveness

Thomas Nind[1], Jeremy Wyatt[2], Ian Ricketts[1], Paul McPate[3] and Joseph Liu[2]

Abstract. Credibility is closely related to trustfulness, reliability, accuracy, authority, bias and quality. There is a strong correlation between credibility of content and its believability [1,2]. Previous studies have identified a number of features impacting on users' assessments of website credibility. In our study, a randomized controlled experiment was carried out with 92 students to investigate the effect of high and low credibility in a website promoting healthy living on the user's behaviour and attitude to exercise. Students allocated to the credible version of the website used it for significantly longer. We believe this demonstrates the importance of designing credible interventions in order to maximise participant exposure. [123]

1 INTRODUCTION

The study of credibility has been identified as an important area of research [3]. Despite the lack of a unified list of dimensions that combine to create credibility (trustworthiness, expertise, competence etc.) there is a general consensus on what features affect perceived credibility.

Credibility is important in all media formats and there is a considerable amount of research on credibility in television, written and spoken communication [4]. The internet however has some unique additional features:

Size - manual assessment of all sites on a specific topic on the internet is not feasible except in the narrowest of fields.

Lack of barriers to publishing - Anyone can create a website at relatively low cost. Authors do not need any qualifications or training in an area before they can publish.

Quality Review - Peer review of internet content prior to publishing is the exception not the norm. Prior to publication a newspaper article will be reviewed by several different editors. Television broadcasts, even when live, are subject to review and those considered to be delivering poor quality or inaccurate information will be withdrawn.

Enforcement - Even if review of internet material was feasible, unless a website contains illegal material, it can not be removed without the author's prior consent.

Validation - Users make judgements about the credibility of a site's content based on the author's credibility. There are a growing number of "phishing" websites that falsify source information in order to exploit visitors [5]. Although there are technical safeguards such as Website Security Certificates and anti-phishing software, current safeguards have been unable to combat this growing problem. New approaches are needed to ensure the integrity of source information.

[1] School of Computing, University of Dundee, DD1 4HN, UK. Email: tnind@computing.dundee.ac.uk.

[2] Health Informatics Centre, Division of Clinical and Population Sciences and Education, University of Dundee

[3] Institute of Sports and Exercise, University of Dundee

There is a strong relationship between credibility and persuasiveness [6]. Pornpitak found that in most situations a highly credible source is more persuasive and that healthcare is an area where additional research on credibility is needed.

We used behaviour change in exercise as a case study. This has been investigated through the Transtheoretical Model (TTM) [7] which allocates people into possible states of change:

pre-contemplation - has no intention to change or denies that there is a need to change.

contemplation - is seriously considering changing their behaviour

preparation - is making small changes to their life to facilitate change

action - is actively engaging in change behaviours such as joining an exercise program

maintenance - continuing to practice change, e.g. exercising

Physical activity declines with age, with the most significant stage of decline in 13-18 year olds [8]. This makes student years an ideal time to target interventions that might restore levels.

Internet and email based interventions have limited impact on behaviour change in physical activity [9]. Improving the quality of these interventions is therefore a high priority particularly with regard to "engaging" and "retaining" participants.

2 WEBSITE CREATION

In keeping with previous studies [10], two versions of a website with the same content and navigation were created. One site was designed with features identified in the literature [1, 11, 12] as improving credibility (Figure 1) and the other with features that erode credibility (Figure 2).

The core content of both websites was the same: background material on exercise and weekly updated articles from reputable sources (Journals, BBC news) identifying it's benefits.

Features associated with credibility included:

Site Awards/Certification - the credible site was accredited by "Health On the Net" (HON). This organisation reviews health-related websites to ensure content is credible and that authors are unbiased and conform to the HON code of conduct. A W3C certification stamp was also included on the credible site to indicate conformance with the XHTML specification ensuring browser compatibility and accessibility.

Photographs of the organization's members - The Assistant Director of the Institute of Sport and Exercise (ISE) consented to have his photograph taken and it was added to the "about us" section of the credible website.

Website Contact Details - The credible site included email addresses and telephone numbers of the website manager and the ISE assistant director at the foot of each page.

Links to external sources & details of the author's credentials for each article - news stories on the credible site included full

references and links to the source materials. No sources were given for material on the control site.

Non-profit - A short message was added to the front page of the credible site informing the user that the "site is non-profit and as such no adverts are displayed or commercial products endorsed".

Content Policy - the credible site included an additional page containing the university's privacy policy.

Display of organization's physical address & contact phone number -The physical location of the university was given in the "about us" page of the credible version along with photographs of the facilities and the telephone number.

Familiar Branding - The credible site used the theme of the university students' sports union with the colour scheme and icons as used by the ISE website.

Interactivity - The credible site contained a rating system where users could vote on the quality of articles and view the average rating.

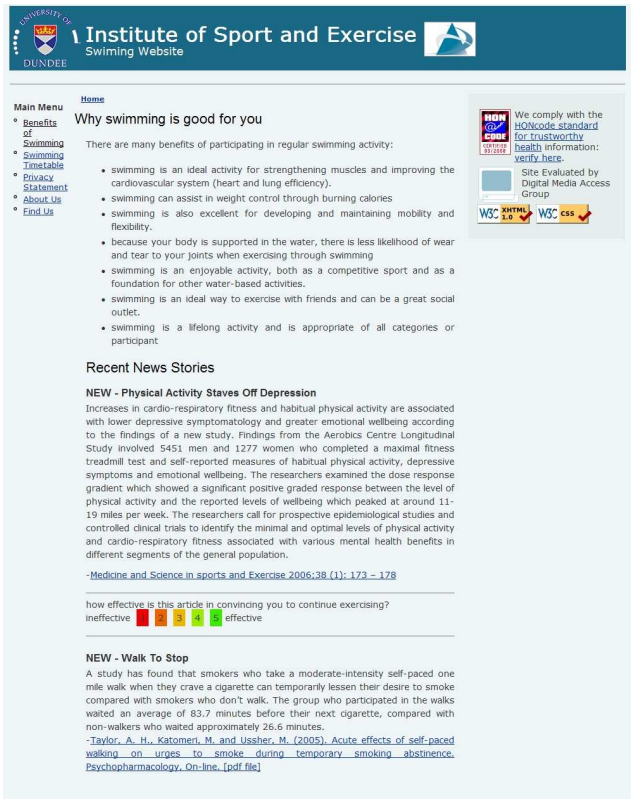

Figure 1. High Credibility site as seen in Internet Explorer at 1024x768 resolution

Reliability - The less credible site contained a broken link on the front page. This link replaced the "about us" link on the credible site to ensure that factors such as menu length and site navigability remained constant. This approach was used rather than more severe methods, such as forcing extended page load times or taking down the site periodically, as this was felt to have too negative an impact on site usability.

Adverts - The less credible site contained a Google advert bar. Google adverts were used because of their familiarity to the user group and prevalence on the web. It was felt that more intrusive advertisements such as "pop-ups" would have too negative an effect on the site's usability.

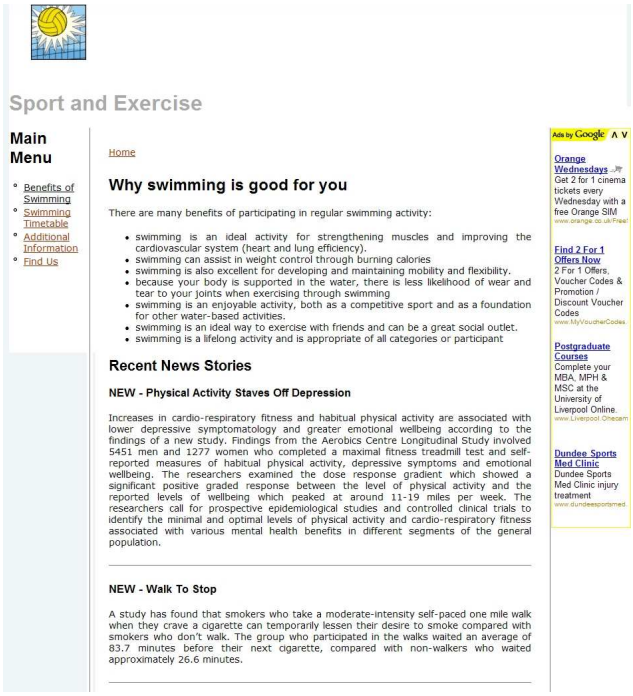

Figure 2. Low Credibility control site as seen in Internet Explorer at 1024x768 resolution

To ensure that both versions of the site were equally accessible, a usability evaluation was performed by a member of the Digital Media Access Group (DMAG) [13]. Both sites were assessed to be equally useable with the exception of the broken link. The reading level of website pages was evaluated using the Simple Measure of Gobbledygook Grading (SMOG) [14]. Both versions of the main website had an average reading age of 12.5, well below the average age of our expected audience. The SMOG grading of the news stories was higher at 15.51 but still well within acceptable limits for a postgraduate student audience.

3 STUDY DESIGN & MEASURES

With the agreement of the University of Dundee ethics committee, a recruitment message was emailed to all current postgraduate students inviting them to participate in a 4 week study investigating attitudes to exercise and the effectiveness of a website promoting exercise.

Participants who followed the hyperlink in the email to the study website were prompted to enter their email address. This was our unique identifier for all participants and also allowed us to send update emails when stories were added to the site to encourage return visits.

After providing their email address and consenting to the study, each participant was randomised to one version of the site and their baseline attitude to exercise was measured using the ISE validated "Physical Activity Preference Questionnaire". Participants were also asked for their age, gender, time spent on

the internet per week and how much of what they read they believed.

Participants' use of the site was logged via a number of server and client side scripts which recorded:

- Time of entry and departure from each page.
- Clicks on the broken link.
- Clicks on the references in news stories.
- Clicks on the "site award" stamps.

New stories were added to the site each week and links emailed out to participants. After 4 weeks an invitation was emailed out to participants to fill in the questionnaire again to measure any change in attitudes to exercise.

4 RECRUITMENT FIGURES & RESULTS

The initial email message was sent to 1584 postgraduate students with a reminder sent 2 weeks later. 233 responded to the email and visited the site. 134 completed the baseline questionnaire and were randomly assigned to one of the two sites. 92 completed the exit questionnaire after using the site for 4 weeks (46 in the control group and 46 in intervention group).

A t-test indicated a significant difference ($p=0.0077$) between the time users spent on the credible & control sites:

- Less credible: mean 54, SD 44, median 46.5 seconds
- Credible: mean 88, SD 71, median 75.5 seconds

The number of pages visited (including repeat visits) also showed a significant difference. The less credible group's mean was 2.6 pages viewed versus 3.7 in the credible group i.e. people allocated to the credible version visited 1 more page.

It is assumed that spending more time on the site and visiting more pages would indicate a more effective and engaging site. It is possible that a shorter visit only indicates a site is better at getting information across but in this case the core content of both sites was the same.

From the baseline and exit questionnaires, the self-reported exercise for each user showed a small but not statistically significant increase ($p=0.0980$) over the 4 week period with the less credible group's mean decreasing from 3.84 to 3.69 and the credible group's mean increasing from 3.85 to 4.0 where the scale was 1: no exercise, 3: not enough to be regular exercise, 5: regular exercise over last 6 months. Other questions regarding attitudes to exercise showed no significant difference.

There was a 10% contamination rate where, over the course of the study, where participants became aware that there were two versions of the website. This was measured after completing the exit questionnaire by asking: "Over the study period, did you become aware that there were other versions of the website?". This was too small a contamination rate to influence study results [15].

5 CONCLUSIONS & STUDY LIMITATIONS

This study demonstrates that a credible intervention is more engaging and is able to hold participants' attention for a longer period of time. However, we were unable to show a change in attitude or in exercise behaviour.

A multi-factorial design would have allowed each individual feature / combination to be evaluated and would have given a more comprehensive view of the relative impact of the manipulations. However we chose to develop only two extreme versions of the site because the first step is to try and demonstrate a difference in exercise uptake between the two most extreme sites before exploring which features lead to this. Also, we had access to a limited number of participants (134 recruited, 93 successfully completed). Running the site as a public resource would allow for a far larger sample size and might help reduce any confounding variables such as biased behaviour due to the fact this was a research study, the Hawthorne Effect. However there are ethical considerations when providing information of low credibility to the public outside of the tight control of a local study.

The size of the website used (6 pages) was also a limitation of this study. A larger site would yield greater precision when looking at the number of unique pages requested. 25% of users explored every page on the website. Initially it was planned to tie participation in the study to each student's attendance at the ISE's facilities for exercise but due to an unplanned switch in database manager software, during the study, this information was lost.

A larger study across all student years is planned along with the development of a longer period of the intervention with more engaging content.

The effect of the credibility factors could only be measured if they were used. For example only 6.5% of participants in the credible group clicked any of the certification stamps but it is possible that other users noted them and were influenced by their presence but didn't follow the link to check their validity. Likewise only 13% of participants actually followed a reference.

When considering changes in behaviour as an outcome measure, it is important to consider the persuasiveness of the core content before manipulations of its credibility. If content is insufficient to motivate a change in attitude or behaviour then manipulations to credibility are unlikely to show a difference.

Evaluating participants' behavioural stage of change prior to recruitment to the study would be beneficial and could be used to screen out pre-contemplators. This would increase the likelihood of seeing a positive impact on physical activity. Alternatively, a stage-matched intervention could be used to deliver appropriate content to each behavioural state group.

REFERENCES

[1] R.S. Young, 'Judgment of information quality and cognitive authority in the web', *Journal of the American Society for Information Science and Technology*, 53, **2**, 145-161, (2002).

[2] B.J. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani and M. Treinen, 'What Makes Web Sites Credible? A Report on a Large Quantitative Study', *Proceedings of the SIGCHI conference on Human factors in computing systems*, 61-68, (2001).

[3] S.Y. Rieh and D.R. Danielson, 'Credibility: A Multidisciplinary Framework', *Annual Review of Information Science and Technology*, **41**, 307-364, (2007).

[4] Poynter Online - Media Credibility Bibliography, Retrieved from http://www.poynter.org/content/content_view.asp?id=1205 on 26/01/09.

[5] P.C. van Oorschot, S. Stubblebine, 'Countering Identity Theft through Digital Uniqueness, Location Cross-Checking, and Funneling', *Financial Cryptography and Data Security* 2005, 31-43, (2005).
[6] C. Pornpitakpan. 'The Persuasiveness of Source Credibility:A Critical Review of Five Decades' Evidence', *Journal of Applied Social Psychology*, 34, **2,** 243-281, (2004).
[7] J. Prochaska and B. Marcus, 'The transtheoretical model: applications to exercise behavior', *Medicine and Science in Sports and Exercise*, 26, **11,** 1400-1404, (1994).
[8] J. F. Sallis, 'Age-related decline in physical activity: a synthesis of human and animal studies', *Medicine and Science in Sports and Exercise*, 32, **9**, 1598-1600, (2000).
[9] A.L. Marshall, N. Owen and A.E. Bauman, *Journal of Science and Medicine in Sport*, 7, **1**, 74-80, (2004).
[10] M.J. Dutta-Bergman, 'The impact of completeness and Web use motivation on the credibility of e-health information.' *Journal of Communication*, **54,** 253-269, (2004).
[11] Z. Liu, 'Perceptions of credibility of scholarly information on the Web', *Information Processing & Management*, 40, **6**, 1027-1038, (2004).
[12] M.J. Metzger, A.J. Flanagin and L. Zwarun, 'College student Web use, perceptions of information credibility, and verification behavior', *Computers & Education*, 41, 271-290, (2003).
[13] Digital Media Access Group...excellent accessibility research and consultancy, Retrieved from http://www.dmag.org.uk/ on 02/03/09
[14] G.H. McLaughlin, 'SMOG Grading - a New Readability Formula', *Journal of Reading*, 12, **8,** 639-646, (1969).
[15] C.P. Friedman, J.C. Wyatt, 'Evaluation methods in biomedical informatics (2nd edition)', *New York: Springer-Publishing*, (2005).

Healthy persuasion: web sites that you can trust

Elizabeth Sillence[1], **Pam Briggs**[1], **Peter R. Harris**[2]

Abstract. Health websites are an abundant and frequently used source of information and advice. How are users of such sites persuaded to trust the advice they read and to act upon it? In this short paper we outline a model of trust in online health advice and highlight the key features in relation to both traditional web 1.0 websites and web 2.0 health sites. Findings from two recent studies are reviewed to provide examples of the ways in which persuasive technology is influencing health behaviour in complex and subtle ways.

1 INTRODUCTION

Health applications are a key target for developers of persuasive technology. Computers, websites and increasingly mobile applications have the potential to change attitudes and behaviours through persuasion and social influence. Encouraging certain positive health behaviours and discouraging negative behaviours has long been the goal of health psychologists, the medical profession and policy makers in general. Health websites offer a way of providing consumers with information and advice about a range of conditions, diseases and lifestyle choices. Who provides this information, the way in which it is presented and the characteristics of the consumer themselves all affect the extent to which the advice is seen as trustworthy and hence the extent to which they are persuaded to act upon it.

In this paper we review the major trust issues associated with e-health and outline a staged model of trust in this domain. This model is then used to highlight two key drivers of persuasive technology: firstly the role of credibility in persuasion and its reliance on design where traditional web 1.0 sites are concerned and secondly the role of people as persuaders in web 2.0 health sites.

2 TRUST ISSUES IN E-HEALTH

Despite its unregulated and often unreliable nature, the Internet is rapidly becoming a new "object of trust" [1]. Research indicates that internet users' rate trust as an important issues within the health domain [2]. They are interested in health advice which is independent and impartial and want websites to be easy to use [3]. Recently health consumers have been turning away from more regulated sites (i.e. those run by government bodies) and towards more personalized sites, often maintained by interested individuals [3] Users are keen to explore other patients experiences online [4] and the rise of sites social networking sites such as myspace and facebook facilitates the disclosure of personal health information. Of course this raises important issues concerning the way in which people evaluate the trustworthiness of health information and advice online and how they choose to engage with health websites.

Various factors appear to be influential in fostering trust. For example, some researchers argue that consumer trust (or a related construct, credibility) is primarily driven by an attractive and professional design [5]. Others argue that trust reflects the perceived competence, integrity predictability and/or benevolence of the site [6] and a few authors also highlight the importance of personalization in the formation of trust judgments [7]. A staged model of trust helps to reconcile the differences in the literature.

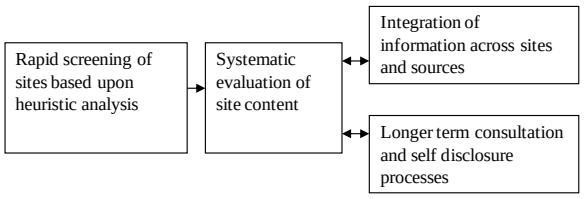

Figure 1. Staged Model of Trust

A number of authors [8, 9] have suggested that three phases are important: a phase of initial trust followed by a more protracted exchange which then may or may not lead to a longer-term trusting relationship. If one considers trust in this developmental context then some of the findings in the literature make more sense. In particular, consideration of a developmental context helps to reconcile the tension between those models of trust which suggest that it is a concept grounded in careful judgment of institution and process factors such as vendor expertise and experience, process predictability, degree of personalization and communication integrity and those models that suggest trust decisions depend much more heavily on the attractiveness and professional feel of a site. A framework built upon one such model [8] proposes three stages of trust development in online health domains (see figure 1). The first stage consists of a heuristic screening process in which sites are rapidly rejected on the basis of their design appeal. This is followed by a second stage in which users undertake a more careful content evaluation of the site noting for example, authorship and credibility issues. The third stage consists of a process of longer term engagement with the site through source integration and self-disclosure processes. This model highlights elements that are important to both traditional web 1.0 health sites and web 2.0 sites. In the remainder of this paper we explore those elements in more detail. Firstly the importance of trust and design cues in changing health behaviour and secondly the role of people as persuaders in an online cancer support group.

3 PERSUASIVE HEALTH WEBSITES

Large scale survey data as well as in-depth qualitative studies suggest that consumers are persuaded by the design elements of health websites [4, 5]. People are more likely to find information and advice on a well designed website credible. Finding

[1] School of Psychology and Sports Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. Email: {elizabeth.sillence, p.briggs}@northumbria.ac.uk

[2] Psychology Department, University of Sheffield, Sheffield, UK. Email: {p.harris}@sheffield.ac.uk

information credible is one thing but being persuaded to act upon the advice is another. Do these design cues actually persuade people to change their health behaviour? In a recent study [10] we assessed the extent to which website design elements or trust cues could influence health behaviour. Participants were shown one of two versions of a website describing the genuine link between alcohol consumption and breast cancer. The Web sites contained the same high-quality content but crucially included design elements or cues known to be either positively or negatively associated with trust. Examples included a TRUSTe seal (positive) and advertising content (negative). Initially participants were persuaded by the material on both websites, One week later, there was a significant interaction between condition and baseline consumption on reported alcohol consumption, the women with higher levels of alcohol consumption who had been presented with positive trust cues reported greater levels of alcohol reduction than those presented with negative trust cues.

4 PERSUASIVE TECHNOLOGY AND WEB 2.0

So in web 1.0 the key to persuasion is credibility. Notions of credibility or trust (in its related active form) are highly dependent on design factors, at least initially. We also know that users are likely to treat computers in a social manner that is to treat the technology as if it was a real person [11]. Is this the case in web 2.0? Web 2.0 is a 2nd generation of web based communities and services such as social networking sites and blogs and aims to facilitate creativity, collaboration and sharing amongst users. The notion of Web 2.0 encompasses a vision of genuine interactivity in which web users are as actively engaged in creating and uploading information as they are downloading and reading web material. Here it appears to be the users themselves that are involved in persuading, rating and recommending so how does this play out in a health context?

People prefer information and advice that is targeted at and written by people like themselves. People viewing health websites are more likely to trust advice that comes from someone based in the same country or from someone who has a similar medical history or lifestyle [4]. Other people can act as powerful instruments of persuasion within health websites. Sillence et al [4] detail how one study participant was persuaded by an online account from a fellow hypertension sufferer to go back to his family doctor to have his medication altered

It also seems that people are far more likely to seek out and be persuaded by very like minded people. In a recent discourse analysis of a cancer support group we explored how participants manage to ask for and offer advice within a peer setting [12]. The support group have developed mechanisms for portraying their competence and trustworthiness and advice seekers seek out very like minded others to provide support for their pre-existing views, thus being more easily persuaded by people with similar views and developing elaborate ways of subtly disregarding information and advice that is not congenial with their way of thinking.

5 SUMMARY AND FUTURE WORK

Persuasive technologies are changing. The migration from the traditional top down approach of web 1.0 to the participant driven web 2.0 health sites has implications for the ways in which people will be persuaded to act upon the information and advice they read. In turn this will influence their health behaviour. A simple heuristic for users is that poor design is indicative of an untrustworthy website. Even the effects of this rule of thumb appear to be more subtle than first thought. Whilst design features can influence responses to health risk information the effects of these trust cues may be slow to manifest themselves influencing health behaviour only with time. With web 2.0 users themselves are the persuaders. They upload content, give advice and make recommendations. The ways in which people respond to health risk information and advice again appears to be complex with information processing styles coming into play. A user with a defensive processing strategy will not be open to persuasion from all sides of the debate. Understanding the ways in which these factors influence responses to information and ultimately health behaviour remains one of the most interesting and important challenges for persuasive technology.

With this in mind we are currently embarking upon a research plan which will explore a key issue associated with Web 2.0 that of patient experience. We already know that some studies have shown that anecdotal, narrative evidence increases perceptions of personal risk and intention to change behaviour [13]. However we still know little about the types of patient experience information that consumers prefer and the ways in which engagement with such material informs their decision making. Over the coming months we will be exploring a number of related questions which aim to increase our understanding of this aspect of Web 2.0 and health advice.

1. What cues do consumers use to guide their searches for patient experience material?
2. Which kinds of patient experience formats do consumers engage with?
3. How do consumers process patient experience material and how can we assess its impact upon decision making? How do they integrate this material with other sources of information e.g. doctors advice, friends and family?
4. To what extent does the processing stance of the consumer influence their preference for patient experience material over Web 1.0 style health advice?
5. What recommendations can we make for the provision and integration of patient experience material in an online environment?

REFERENCES

[1] Hall, M., Zheng, B., Dugan, E., Camacho, F., Kidd, K. E., Mishra, A., & Balkrishnan, R. Measuring patients trust in their primary care providers. *Medical care research and review, 6 (3) e36.(2002).*

[2] Williams, P., Huntington, P., & Nicholas, D. Health information on the Internet: A qualitative study of NHS direct online users. *Aslib Proceedings, 55 (5/6), 304-312. (2003).*

[1] School of Psychology and Sports Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. Email: {elizabeth.sillence, p.briggs}@northumbria.ac.uk

[2] Psychology Department, University of Sheffield, Sheffield, UK. Email: {p.harris}@sheffield.ac.uk

[3] Sillence, E., Briggs, P. Harris, P, Fishwick, L.Going online for health advice: Changes in usage and trust practices over the last five years. *Interacting with computers, 19, 397-406.(2007).*

[4] Sillence, E., Briggs, P. Harris, P, Fishwick, L. Health websites that people can trust - the case of Hypertension. *Interacting with Computers , Special Issue on 'Moving Face-to-Face Communication to Web-based Systems', 19, 32-42.(2007).*

[5] Fogg, B. J., Kameda, T., Boyd, J., Marchall, J., Sethi,R., Sockol, M. and Trowbridge, T.. *Stanford-Makovsky Web Credibiltiy Study 2002: Investigating what makes Web sites credible today,* A Research Report by the Stanford Persuasive Technology Lab & Makovsky & Company, Stanford University. Retrieved from : http://www.webcredibility.org

[6] Bhattacherjee, A. Individual Trust in Online Firms: Scale development and Initial Test. *Journal of Management Information Systems, 19, 211-241.(2002).*

[7] Briggs, P., de Angeli, A. & Simpson, B.. Personalisation and Trust: A Reciprocal Relationship? In M.C. Karat, J. Blom and J. Karat (Eds). *Designing Personalized User Experiences for ECommerce*. *Kluwer.(2004).*

[8] Sillence, E., Briggs, P. Harris, P, Fishwick, L. A framework for understanding trust factors in web based health advice. *International Journal of Human Computer Studies, 64, 697-713. (2006)*

[9] Egger, F. N."Trust Me, I'm an Online Vendor": Towards a Model of Trust for E-Commerce System Design. *Proc. CHI 2000*. ACM Press.*(2000).*

[10] Harris, P. R., Sillence, E. & Briggs, P. The effect of credibility related design cues on responses to a web based message about the effects of breast cancer risks from alcohol. (under review).

[11] Reeves, B. & Nass, C. *The media equation: How people treat computers, television and new media like real people and places.* Stanford, CA: CSLI Publications (1996).

[12] Sillence, E. Seeking out very likeminded others: Exploring trust and advice issues in an online support group. *International Journal Web Based Communities* (in press).

[13] De Wit, J. B. F., Das, E. & Vet, R. What Works Best: Objective Statistics or a Personal Testimonial? An Assessment of the Persuasive Effects of Different Types of Message Evidence on Risk Perception. *Health Psychology, Vol. 27, No. 1, 110–115 (2008).*

[1] School of Psychology and Sports Sciences, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK. Email: {elizabeth.sillence, p.briggs}**@northumbria.ac.uk**

[2] Psychology Department, University of Sheffield, Sheffield, UK. Email: {p.harris}**@sheffield.ac.uk**

Adaptive Persuasive Scripts

Maria E. Pertou and Henrik Schärfe[1]

Abstract. In the context of the HANDS project [13], we argue that cognitive support systems designed for mobile devices may benefit from classical AI techniques as well as from Persuasive Technology. We investigate a system of stepwise instruction, called a Simple Safe Success Instructor, first in terms of *scripts*, and later in terms of *Hierarchical Task Networks.* The system under development is crafted for the benefit of young people with an autism-diagnosis, and the purpose is to support these in desired changes in attitude and behavior in dynamically changing environments.

1 INTRODUCTION

In the pursuit of designing persuasive technology we find it fruitful to look into the findings of AI; since when combined with the theories of persuasive technology (PT) it will be possible to design persuasive technologies that are a lot more adaptive and therefore also potentially more persuasive than conventional PT. An example of a persuasive technology that will gain from an outlook to research in AI is the Simple Safe Success Instructor (SSSI), which is a part of the HANDS project. The HANDS project aims at helping young people with an autism-diagnosis navigate and develop socially through persuasive technology. The general idea is to design a computer program that allows teachers and key caretakers to build cognitive support systems of various kinds. These support systems are then uploaded to the student's PDA. The project is in nature highly interdisciplinary, and the consortium conducting the research consists of experts in pedagogical design for children with autism from LSBU in London; experts in autism-spectrum disorders from the point of view of cognitive psychology from ELTE University in Budapest; and experts in persuasive design from Aalborg University, Denmark.

The actual prototypes are being developed based on input from the three universities in close cooperation with teachers and students from four schools in Budapest, Meopham, Kent, Stockholm, and Aalborg. The target group for the HANDS system consists of teenagers of normal intelligence, but with issues often related to diagnoses within the autism spectrum. This means that the students typically are high-functioning in many respects, but may also experience problems with handling situations outside a strict regime of routines [14].

In the early stages of the project, and leading all the way up to the software specifications, user scenarios have played an important role. In this paper, we will focus on one such scenario, and we will consider various properties of cognitive support systems from a PT point of view. In section 2, we suggest to draw on the AI tradition of scripts and frames in order to formalize the problem. In section 3, we suggest to model a user scenario in terms of hierarchical task networks.

.

[1]Dept of Communication and Psychology, Aalborg University, Denmark. Email: {marip, scharfe}@hum.aau.dk

In section 4, we relate these structures to known principles in Persuasive Technology. Finally, in section 5, we investigate the possibilities of extending the case scenario with location-aware capabilities.

2 SOCIAL STORIES AS SCRIPTS

The software requirements for the HANDS toolset is written in use case format, in which scenarios and user stories play an important role [3, 15]. The user stories enable experts in non-technical domains to specify success criteria for the software. In many cases this consists of a process where different persuasive needs are clarified. The needs may differ from child to child and from school to school, but the scenarios are still deemed typical beyond the domain of one child and one school. A scenario often considered in the development of the HANDS toolset is the case in which a student has to travel on his own using public transportation. This situation is realistic in the sense that it is a task many people in the target group wants to be able to do, and also finds difficult because of the number of things that might happen, and can possibly go wrong. Even small things that do normally not pose problems for neuro-typical teens, may here cause severe frustration and failure to complete the tasks in question. In many schools for autism-diagnosed children and youths, it is customary to use *social stories* to assist the user in keeping track of step-wise instructions as they unfold. In addition, social stories are also used for rehearsing difficult situations [6].

Social stories are not necessarily stories in the sense that they embody narrative qualities, but rather they are step-wise accounts of what typically happens in certain situations, or they contain heuristics for desired behavior in given situations. In this sense, we might say that the narratives embody meaning central to acting in the domain of the users, and also as an important instrument for the software developers [10]. Specifically, the software must support the teachers in writing cognitive support systems that are tailored to the needs of individual children, aesthetically pleasing, and which embody effective means of dealing with situations, that the child in question finds difficult.

While collecting the use cases and user stories, we have found that a number of scenarios appear repeatedly, and we consider these stories to be powerful instruments in shaping expectations among and between future users of the system. At the same time, teachers and other caretakers insist on the importance of being able to customize cognitive support systems directly to an individual user. In fact, we have repeatedly heard practitioners say that the support systems must be manufactured precisely to one child, and that the software designed for the teachers therefore must facilitate this extreme level of customization. Although social stories are highly adapted to individual users, we argue that these cognitive support systems can be said to have a logical core in two respects. In the first place, many teachers do in fact report that they deal with many of the same

situations, typically situations such as morning routines, going on a bus, meeting people, buying groceries and similar everyday practices. So even though the actualized story has to be highly individualized in order to fit the student's needs, the stories, at the same, time refer to essentially typical situations, and are recognizable as such. In the second place, the stories themselves are in many cases structured around a logical core with a rigid, formalizable structure. In artificial intelligence research we have a long tradition for talking about such structures as scripts. In the words of Schank and Abelson: "Some episodes are reminiscent of others. As an economy measure in the storage of episodes, when enough of them are alike they are remembered in terms of a standardized generalized episode which we will call a script" [11].

A script, e.g., for taking the bus, will vary according to cultural and local differences, but most people will as a minimum recognize that the script includes buying a ticket, locating the right bus at location A, entering the bus at location A, riding the bus, and exiting the bus at location B.

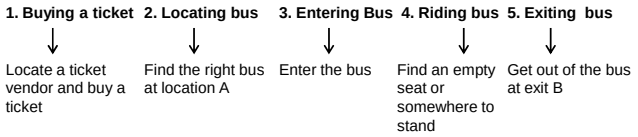

Figure 1. Script for taking the bus

For the teachers working with autism-diagnosed youths, writing such scripts is a complicated task that involves a very deep understanding of the individual person. This process normally involves detailed information stemming from a long-term relationship between a teacher and a child. The question becomes to deeply understand how the individual construes a given situation, for instance figuring out exactly which part of an everyday situation that may cause problems. A fundamental design idea underpinning the HANDS toolset is therefore a desire to build interfaces that not only allow the teacher to compose files containing SSSIs or social stories in an appropriate and aesthetically pleasing manner, but the teachers should also have tools that allow them to evaluate the persuasive effect of the HANDS enabled PDAs. We are, in other words, interested in supporting existing practices by means of ICT; and also to fertilize existing practices by means of advanced conceptual modeling and advanced feedback systems.

In terms of persuasion this means that the HANDS toolset has a double scope of intention. We are, in the first place, hoping to give teachers tools that will aid them in monitoring the use of the cognitive support systems they produce in novel ways, thus creating a room for reflection on practice. The more direct persuasive objective is of course to support the children and youths in navigating more freely in, what appears as very complicated settings for them. Other efforts as well as pilot-studies indicate that persuasive technology may be a significant contributor to this.

3 FORMALIZING SCRIPTS IN HTN

We propose to further formalize the scripts by means of a Hierarchical Task Network (HTN). There are two reasons for this choice. In the first place, the AI tradition dating back to [9, 12] of dividing a task into sub-tasks has proven to be a very efficient tool for automated planning, see also [7]. Specifically, we follow the tradition of using AND-OR graphs to represent several sub-routines that must all be satisfied in order to solve a given problem [8]. Seeing the situation as a script provides us with a good understanding of the subtasks in question. In the second place, HTNs have in recent years proven useful in modeling dynamic contexts [2], and furthermore, HTNs have been proven to function as a central component in interactive systems with a high degree of user influence [1].

Figure 2 shows a HTN structured as an AND-OR graph where the overall goal is to use public transportation to get from location A to location B. In this example the public transportation is in form of a bus ride and the sub goals that the overall goal is divided into are listed at the level below. This level consists of tasks that must be fulfilled in order to accomplish the overall goal. The tasks may vary according to cultural and local settings, and must therefore be adjusted to match those differences. The arcs are read as "AND" and represent the tasks necessary to complete the overall task. The edges with no arcs are read "OR", and represent situations where the task has more than one possible solution. It is, e.g., possible to either take a seat or remain standing during a bus ride.

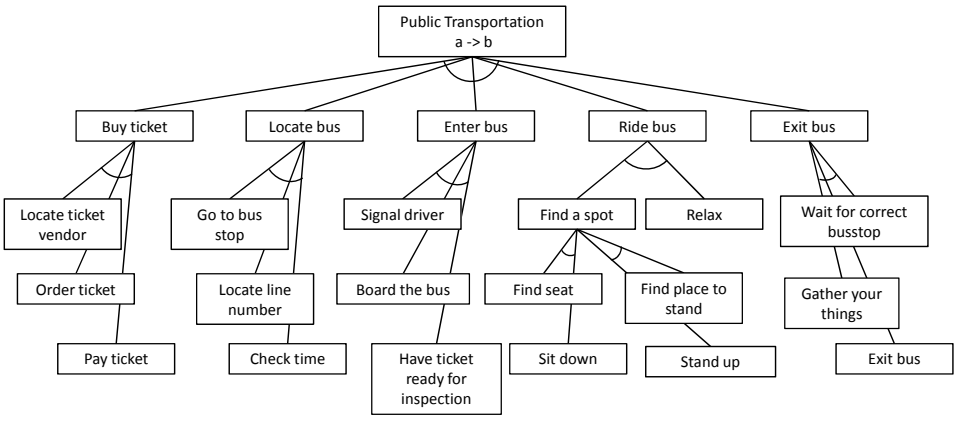

Figure 2. HTN for public transportation

The student should do one or the other – not both. The whole series of events can also be presented using branching time to illustrate the possible future courses of events [14]. This is however beyond the scope of this paper.

4 EMPLOYING PERSUASIVE PRINCIPLES

The AI principles of scripts and HTN are not necessarily persuasive in themselves, but if they are combined with persuasive principles, we believe that they can prove very helpful as extensions or improvements of some of the persuasive strategies put forward by B.J Fogg [4].

The main challenge for the SSSI is to be able to define the most opportune moment for the system to intervene. In the ancient theories of persuasion, this is framed as Kairos [5]. Fogg refers to this as *suggestion*. If the script is subdivided into tasks that the user has to go through in order to fulfill the overall goal, say, that of taking the bus, it becomes more obvious when the system should intervene. Particularly, this is true of situations where actual circumstances do not match those outlined by the system. If the bus is delayed or does not show up at all, the system cannot proceed and must return to earlier steps that can push information on how to find another bus.

It is a general challenge for PT to keep the user interested in the technology and the actions it suggests. The *tunneling* strategy, suggested by Fogg forms a possible solution to this problem. However, when the technology is used in a real-life situation it becomes even more challenging to keep the user interested since the user is not tied to one specific location, e.g., in front of the computer. It is therefore not enough for the system to simply suggest actions in a predetermined order. The system must also be highly adaptable to possible changes in the situation.

When the system is equipped to gather contextual information, it also becomes possible to present that information in an authoritative manner. This opens the possibility of letting the system take on the form of a *social actor*; yet another strategy proposed by Fogg in order to influence behavior. In fact, the systemic interaction between the SSSI and other parts of the HANDS toolset, gives rise to many other persuasive strategies as well, including giving rewards, and various kinds of monitoring.

5 EXTENDING THE SSSI

Simply having the script present on a mobile device may be of great help in many cases. The advantages of the HANDS toolset here consist in having access to familiar information in a customized form, and in a non-obtrusive format (on the screen of a mobile phone rather than, say, on laminated cardboard pieces). But several extensions may be added to this skeleton version, which may enhance functionalities, user experience, and persuasive potential. Within a framework that can be formalized through HTNs, we envision the following levels that may be added to the basic functionality.

1. *Adding static information about routes and locations.* Firstly, the public transportation SSSI may be extended by integrating timetables for trains and busses.
2. *Adding static information about the user.* Secondly, timetables may be integrated with actual information regarding the user's destinations; be it actual destinations extracted from the user's calendar, or possible destinations, i.e., extracted from the user's list of contacts and locations. Since time planning is essential to any work with autism-diagnosed youths, the HANDS toolset also has strong calendar functions build into it [15]. For example, integration of addresses from a list of contacts is found in the TOMTOM Navigator, which can synchronize addresses with MS Outlook. Integration with calendars and and contact lists help adding routes and times to the cognitive support system.
3. *Adding dynamic information about the context.* Thirdly, the system may be enriched by contextual information at runtime. A typical problem arises when a proposed plan is disturbed, e.g., if the bus is delayed. GPS technology can be used to convey such information to users, possibly hindering a breakdown in this case. Figure 3 shows a mobile interface indicating which busses will arrive next at a given bus stop. The + sign indicates that the first bus to arrive is 1 minute late. For people with autism-diagnosis, even slight delays can be very disturbing. A user waiting for the second bus in figure 3 might benefit from a discreet prompt: "The bus is 3 minutes delayed. You can still make your connecting bus".

Figure 4. Possible extensions

4. *Adding dynamic information about the context of the user.* Fourthly, the user's location can be logged via GPS, allowing the system to detect discrepancies between the plan and actual events. This gives the possibility of considering intervention at the appropriate moment. Let us say, for instance, that a user for some reason misses a bus, corresponding to the facts that the user is at the expected bus stop (cf. 2), but too late (cf. 1), and the bus has already left (cf. 3). For some users, the toolset should now intervene and prevent panic by prompting the user: "The bus already left. Want to look for the next solution?"

Figure 4, below, shows which tasks may gain from being extended by static and dynamic information. The higher the number of the extension the more complex it is. The complexity increases the chances that the system, if it works as intended, will be persuasive and able to intervene at the most appropriate time.

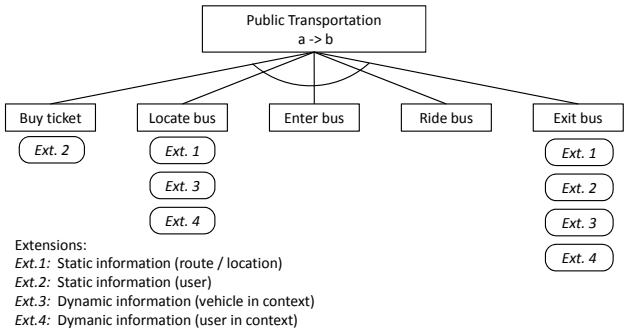

Figure 4. Possible extensions

6 CONCLUSION

In this paper we have suggested to draw on the AI tradition of scripts in order to formalize the social stories that are used in order to help young people with an autism-diagnosis handle difficult situations. The formalization makes it possible to divide the script into subtasks that can be implemented in a HTN. Certain extensions may be added to the SSSI in order to make it more adaptive, namely static and dynamic information about routes, locations, the user, and the context. Furthermore, we have considered the relations between AI modeling and persuasive potential.

ACKNOWLEDGMENTS

The authors acknowledge the stimulus and support of the EC-sponsored project, 'Helping Autism-Diagnosed teenagers Navigate and Develop Socially (HANDS)', sponsored by the European Commission as part of the 7th Framework Programme, Theme 7.2, Accessible and Inclusive ICT.

REFERENCES

1. Cavazza, M., et al. *A 'Companion' ECA with Planning and Activity Modelling*. in *International Conference on Autonomous Agents*. 2008. Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems.
2. Charles, F., et al., *Compelling Experiences in Mixed Reality Interactive Storytelling*, in *International Conference on Advances in Computer Entertainment Technology, ACE '04*. 2004: Singapore.
3. Cockburn, A., *Writing Effective Use Cases*. 2000: Addison Wesley.
4. Fogg, B.J., *Persuasive Technology - Using Computers to Change What We Think and Do*. 2003, San Francisco: Norman Kaufmann Publishers.
5. Glud, L.N. and J.L. Jespersen. *Conceptual Analysis of Kairos for Location-based Mobile Services*. in *Persuasive 2008. The Third International Conference on Persuasive Technology*. Oulu, Finland: Oulu University Press.
6. Gray, C., *My Social Stories Book*. 2001: Jessica Kingsley Publishers.
7. Nau, D.S., S. Smith, J. J., and K. Erol, *Control Strategies in HTN Planning: Theory Versus Practice*, in *American Association for Artificial Intelligence*. 1998, AAAI/IAAI-98 Proceedings: Madison Wisconsin, USA. p. 1127-1133.
8. Rich, E. and K. Knight, *Artificial Intelligence. Second Edition*. International Edition ed. 1991: McGraw-Hill, Inc.
9. Sacerdoti, E.D., *Planning in a Hierarchy of Abstraction Spaces*, in *IJCAI 1973*. 1973: Stanford, California, USA. p. 412-422.
10. Schank, R., *Tell Me a Story. Narrative and Intelligence*. 1990: Northwestern University Press.
11. Schank, R. and R. Abelson, *Scripts Plans Goals and Understanding An Inquiry into Human Knowledge Structures*. First ed. The Artificial Intelligence Series. 1977, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers. 249.
12. Tate, A., *Generating Project Networks*, in *IJCAI-77*. 1977: Cambridge, Massachusetts, USA.
13. Øhrstrøm, P., et al., *HANDS - Helping Autism-diagnosed teenagers Navigate and Develop Socially. EU project under the 7th Framework Programme, Theme 7.2, Accessible and Inclusive ICT*. 2008, http://hands-project.eu.
14. Øhrstrøm, P., H. Schärfe, and M. Gyori, *A Conceptual Analysis of Difficult Situations – developing systems for teenagers with ASD*. 2009 (forthcoming).
15. Aagaard, M., et al., *HANDS Deliverable D4.2.1 Requirement Report http://hands-project.eu/*. 2008, AAU.

Constructing a Rhetorical Figuration Ontology

Randy Harris and **Chrysanne DiMarco**[1]

Abstract. Many essential components of language charted by rhetoric, the ancient study of persuasion, remain understudied and underrepresented in current Natural Language systems. Our goal is to combine linguistic and rhetorical theories with discourse analysis and machine learning to develop formal models of computational rhetoric that may be usefully applied in real-world Computational Linguistics systems. As part of this initiative, we are building an ontology of rhetorical figures and formalizing their expression.

1 INTRODUCTION

Natural Language Processing (NLP) systems are now sufficiently advanced for use in everyday business, educational, and personal applications. Search engines, a prime example, have become an essential part of how academics do their research, how businesses follow trends, and how the average person accesses the Internet. But NLP systems are still challenged by basic problems in understanding the full significance of a text. Current systems generally deal with only restricted language or use simplified methods of analysis, such as shallow parsing. Too much attention has been placed on semantics at the expense of rhetoric (including stylistics, pragmatics, and sentiment). While computational approaches to language have occasionally deployed the word "rhetoric", even in quite central ways (such as Mann and Thompson's Rhetorical Structure Theory [13]), the deep resources of the millenia-long research tradition of rhetoric have only been tapped to a vanishingly small degree. This tradition studies three general attributes of texts that we can formalize and therefore utilize: style (including lexical choice, syntactic structure, and modes of address); purpose (such as description, persuasion, and instruction); and affect (such as trust, deference, and anger).

Our method of "stylistic patterning" is based on the idea that the meaning of an utterance is communicated through the relations among its constituents, as well as their relations with contextual and co-textual elements. We take the notion of style broadly, as "all the choices a writer [or speaker] makes in his or her words and their arrangement" [11, page 14]. Stylistic patterns may be unintended, but they are never meaningless. The question is, how can we get at stylistic meaning? Our model does it with a range of stylistic and rhetorical effects and functions at various levels of discourse organization. In the context of the discourse, stylistic choices contribute to various rhetorical elements: situational parameters or formality levels, as well as intentions, stances, social identities, or moods [17]. As well, some of the stylistic phenomena collectively referred to as "register" are known to be related to text structure and genre. In this paper, we present the first stages of our approach to building a facility for incorporating the stylistic aspects of rhetoric in computational Natural Language systems, specifically, we describe an *ontology of rhetorical figuration* for describing rhetorical patterns that can be used to create persuasive effects in discourse.

[1] University of Waterloo, Waterloo, Ontario, Canada. Email: raha@uwaterloo.ca; cdimarco@uwaterloo.ca

2 RELATED WORK

Rhetorical theory has become an increasingly valuable resource to researchers in Natural Language Processing as formal grounding for their computational models and systems. Crosswhite [2] puts succinctly the value of rhetorical argumentation theory to computational linguists: the repositories of formal argumentative schemata (e.g., [14]), rhetorical figures (cf. [6] for evidence of their value), and the representation of the audience, both "a particular audience (with particular values and beliefs) and a universal audience (one that is constructed by imagining away the peculiar local beliefs and attitudes of some actual audience and imagining into this audience the requisite intelligence, memory, attention, knowledge, and so on, so that the resulting audience embodies one's concept of rationality)". The second and third issues are of particular interest to us.

Within the computational community, various formal models of rhetoric have been employed. The recent series of workshops on Computational Models of Natural Argument (CMNA) attest to the growing adoption of models such as Toulmin's [18] logical model, Perelman and Olbrecht-Tyteca's [14] argumentation schemes, and Walton's [19] informal logic for analyzing and evaluating natural argumentation. The usefulness of rhetorical argumentation has so far largely been addressed at a rather abstract level of discourse representation, i.e., formal frameworks and schemata of rhetorical argumentation [8, 12, 9]. Our work is more concrete in its aim: to develop computational representations of fine-grained aspects of style and rhetoric, and apply these to problems that require linguistic expressivity, but where computational efficiency is also a key issue.

3 OUR APPROACH

We are combining Computational Linguistics and Rhetorical Theory to develop formal computational models of style, pragmatics, and sentiment that may be applied in Natural Language systems. For instance, a central concern of rhetoric has been stylistic flaws (such as excessive repetition of terms and dysfluencies of reference), as well as stylistic merits (such as clarity and cohesion). With rhetorical diagnostics we can locate textual deficiencies; with rhetorical strategies, we can repair them. With other diagnostics and strategies we can detect and reinforce or alter the purposes and/or the sentiments of texts. We can re-engineer textual elements to clarify purposes, enhance or reduce emotional effects, and therefore reshape texts as a function of audience, genre, and context. In short, we can tailor texts to specific readers and specific needs. This capacity is especially important in an environment with vast textual reservoirs and widely discrepant

audiences, such as health care. Computational rhetoric will, for example, allow doctors and other medical personnel to generate patient-specific brochures, matched to criteria like gender, age, reading-level, symptomology, prognosis, collateral conditions, contra-indications, medication, and so on. Our HealthDoc Project [4, 5] has this aim of automatically generating health education tailored to a patient's individual characteristics and medical condition.

It is a difficult challenge to develop expressive, fine-grained ontologies that will lend themselves to use in computationally efficient textual analyzers. Our approach is based on adaptation of well-established rhetorical theories, representation of this theoretical information in precise computational formalisms (e.g., feature-based logics), and implementation in software systems that are computationally efficient. As first steps, we have developed a rich *stylistic ontology* of finely-detailed linguistic features at multiple levels of description [3] and we are now implementing an efficient *stylistic annotator* based on this ontology [16]. Our approach is based on the belief that surface analysis of fine-grained stylistic features, easier and more tractable than full-scale deep-semantic analysis, can yield significantly more, and more meaningful, information than current shallow parsers or statistical methods in a computationally efficient manner.

As part of this work, we have developed a *rhetorical-figure annotation tool* [7], based on traditional definitions of figures, to manually annotate rhetorical figures in various text corpora (e.g., political speeches, health educational materials). Our next step in this work, the automated annotation of rhetorical figures will however require a formal description—an ontology of figuration, in effect—that can be used to characterize and classify rhetorical patterning for use in tasks like recognizing rhetorical strategies such as persuasion and argumentation; detecting and then "repairing" stylistic dysfluencies such as repetitive or awkward text; enhancing or recalibrating redundancies and saliencies for specific aspects of the message; improving credibility and shaping emotional response; and so on. Towards these ends, we are formalizing a set of rhetorical figures in a manner that lends itself to computational representation.

4 RHETORIC

Since the Enlightenment, heavily abetted by the rise of science and technology and their concomitant theories of language as a neutral, context-free, affectless, transparent vessel of communication, rhetoric has fallen into disrepute, with rhetorical figures prominently targetted as the devices of purposive, context-laden, emotional, and opaque language. "Who can behold, without indignation," Thomas Sprat asked, mounting an early modern assault on rhetoric in his *History of the Royal Society*, "how many mists and uncertainties, these specious Tropes and Figures have brought on our Knowledg[e]?")[2]. We pass over the matter of how accurate the windowpane theory of language is, and even the spectacle of Sprat using a rhetorical question to launch his condemnation of rhetoric, to notice simply (1) that the purposive, contextual, and emotional aspects of language are precisely the ones that interest us, (2) that all language manifests these aspects, (3) that rhetoric is not so much a cause of opacity as a methodology for understanding it, and (4) that Sprat is certainly right to implicate figuration in all of these matters. So, we are building an ontology to make rhetorical figures more tractable computationally.

[2] Thomas Sprat. 1667. *The History of the Royal Society of London, for the improving of natural knowledge*. London, J. Martyn and J. Allestry, Printers to the Royal Society. Page 112.

5 FIGURES

Rhetoricians have been studying figuration for millennia, in many languages, under many different theoretical allegiances, with the result that hundreds of overlapping, inconsistent, and even contradictory taxonomies exist. But this very extensive research has produced a rich basis from which we can develop our ontology, and which we have augmented by work in computational linguistics and cognitive science. At the first level of analysis, for instance, two traditional categories are indispensible: tropes and schemes. Tropes, such as metaphors and synecdoches, are conceptual in nature. Schemes, such as alliteration and polyptoton, are formal.

Metaphors rely on the cognitive principle of comparison (*Jeff is a brick* compares a person to an object known for its solidity), synecdoche on *PartOf* representation (*All hands on deck* identifies sailors by singling out aspects of their anatomy critical for sailorly tasks). The form a trope takes is secondary to the conceptual principles at work. If there is no expression of comparison, there is no metaphor. We can say *A brick, that's Jeff.* We can say, *Jeff is the brick of that family.* And so on, in as many syntactico-lexical configurations as we have the imagination, and we still have a metaphor. But, if we say, *Jeff is a stable guy*, the metaphor is gone (strictly speaking, of course, we bring in another metaphor, since *stable* is fundamentally a physical term, and here we are using it for emotional and social purposes, but it is a subtler, more 'literal' sort of metaphor). Schemes, conversely, are formal in nature, and their conceptual operation is secondary to their structural arrangement.

Alliteration is the consecutive use of words with the same initial consonants (*Peter Piper picked a peck...*); the semantics of those words is irrelevant to the scheme. Polyptoton is the use of one word stem in a variety of morphological instantiations (*That team is the suckingest bunch of sucks that ever sucked*); again, semantics are not part of the equation. For this reason, schemes are of course the most amenable to computational detection and manipulation, and we are concentrating our early energies on them.

6 TOWARD AN ONTOLOGY OF RHETORICAL FIGURES

Despite the extensive number of rhetorical figures that have been catalogued over two millennia[3], they fall into a relatively few, partially overlapping classes. While we have not worked out an exhaustive set of classes and relations, we are especially intrigued by the way in which the natural organizing principles of figures manifest well-known cognitive affinities, like comparison, contrast, and symmetry, and by the interplay of well-known linguistic operations in the patterning of figures, like addition, deletion, and permutation. As an example, consider some schemes of omission, in which normally expected elements are implied rather than stated. The most familiar such figure is ellipsis, in which a lexeme is omitted (e.g., *John forgives Mary and Mary, John*). Ellipsis is a clear example simultaneously of how linguistics as a field has drawn on rhetoric (frequently without acknowledgement or even awareness), and of how figures permeate mundane language processes [10]. But there is a wide range of omission schemes. Zeugma, for instance, is a scheme in which one lexeme (usually the main verb of a sentence, sometimes a noun or adjective) governs two or more other lexemes in a series. Zeugma *IsA* ellipsis, that is. Moreover, zeugma includes multiple types, depend-

[3] Sylva Rhetoricae [http://humanities.byu.edu/rhetoric/silva.htm], a superb online resource for rhetorical figures, lists 433 distinct figural terms.

ing on the kind of governing lexeme or its placement in the clause, including:

Prozeugma: The verb in the first of a series of clauses governs the noun phrases in the remaining clauses in the series (e.g., *Her beauty pierced mine eye, her speech mine woeful heart, her presence all the powers of my discourse.*).

Hypozeugma: A verb follows a series of words or phrases that it governs (e.g., *Friends, Romans, countrymen, lend me your ears.*).

Epizeugma: The verb that completes a predicate occurs at either the very beginning or the very ending of its sentence (e.g., *Fades beauty with disease or age. Either with disease or age beauty fades.*).

Mesozeugma: The verb governing multiple subjects occurs in the middle of a construction that contains them all (e.g., *Neither his father nor his mother could persuade him; neither his friends nor his kinsmen.*).

Each of these figures, in short, *IsA* zeugma (in this case, the relationship is conveniently signalled by traditional nomenclature). Function words might also be omitted: asyndeton *IsA* ellipsis in which conjunctions between clauses are omitted (e.g., *government of the people, by the people, for the people*). At the sub-lexical level, syncope (another clear example simultaneously of how linguistics has drawn on rhetoric and on how figures deeply interpenetrate with ordinary-language processes) *IsA* medial ellipsis of phones or syllables, such as when *library* is pronounced *libary*. Apocope *IsA* terminal ellipsis, in which the final sound or syllable is omitted, as in the back-formation of *pea* from *pease*. A poetic example that includes both apocope and syncope, in that order, is Alexander Pope's *What oft was thought, but ne'er so well expressed*[4].

Figures are also related through opposition. Diazeugma, for instance, *IsOppositeOf* zeugma, since it is a scheme in which a single subject governs several verb phrases (usually arranged in parallel fashion and expressing a similar idea), as in *The Romans destroyed Numantia, razed Carthage, obliterated Corinth, overthrew Fregellae*. Polysyndeton *IsOppositeOf* asyndeton, since it involves the elaborate use of conjunctions between clauses, as in this passage from Hemingway's *After the Storm*:

> I said, "Who killed him?" and he said, "I don't know who killed him but he's dead all right," and it was dark and there was water standing in the street and no lights and windows broke and boats all up in the town and trees blown down and everything all blown and I got a skiff and went out and found my boat where I had her inside Mango Key and she was all right only she was full of water[5].

Figures are also related through relations of inclusion. Consider schemes of iteration. Ploche is simple lexical repetition (*She is tall, very tall*), so it naturally participates in schemes of complex lexical repetition, such as antimetabole (lexical repetitions in reverse order: *Ask not what your country can do for you—ask what you can do for your country*) and epistrophe (repetition at the ends of phrases: *government of the people, by the people, for the people*).

[4] Alexander Pope. *An Essay on Criticism*. Whitefish, MT: Kessinger Publishing, 12, 2004.

[5] Ernest Hemingway. *The Complete Short Stories of Ernest Hemingway*. New York: Simon and Schuster, 283, 1998.

7 FORMALIZING SCHEMES

In our formalization of schemes we use the descriptive elements shown in Table 1. A portion of our current set of rhetorical figures and their formalizations is shown below, together with sample realizations.

Table 1. Formalism for Representing Rhetorical Figures

Element	Meaning
Cl	clause
Phr	phrase
W	word
Vb	verb
N	noun
S	stem
M	morpheme
C	consonant
V	vowel
P	phone
Sy	syllable
$\oslash$	gap
X/Y	X INSTEAD OF Y
$\ldots$	arbitrary intervening material (possibly null, with some upper limit, shorthand is *proximal*)
$\{\ldots\}$	morpheme boundaries
$[\ldots]$	word boundaries
$<\ldots>$	phrase or clause boundaries (assuming clauses are just special types of phrases, aggregating other phrases)
Subscripts	identity (same subscripts), nonidentity (different subscripts)

Adage (apothegm, gnome, paroemia, proverb, sententia, maxim): Use of familiar, traditional expressions.

(1) $X,\ where\ X\ =\ \{x_1, x_2, \ldots x_n\}$

That is, locate any occurrence of a pattern that is stored in an adage box somewhere (*A bird in the hand is worth two in the bush*, *Never look a gift horse in the mouth*, etc.). We are not yet certain how feasible doing this type of recognition automatically would be. There are two complications that we can foresee: first, finding a collection of adages/proverbs/idioms that we could represent easily in the right computational format for the pattern-recognizer; second, specifying substrings appropriately. For instance, we might encounter something like *She preferred a bird in the hand to a speculative treatment*, which evokes the adage, but doesn't fully replicate it.

Alliteration: The repetition of consonants at the beginning of proximal words.

(2) $[C_a \ldots] \ldots [C_a \ldots]$ *Lopsided loons lull listening lovers.*

Anadiplosis: Starting a clause or phrase with the word or phrase that ended the preceding unit.

(3) $< \ldots [W]_a >< [W]_a \ldots >$ *Drake covets loons, loons with cash.*

(4) $< \ldots < \ldots >_a >< < \ldots >_a \ldots >$ *Drake covets lopsided loons, lopsided loons with cash.*

Anaphora: The repetition of a word or group of words at the beginning of successive clauses or phrases.

(5) $< [W]_a \ldots >< [W]_a \ldots >$ *Drake covets loons. Drake loves cash. Drake wants fame.*

(6) $< [W]_a[W]_b \ldots >< [W]_a[W]_b \ldots >$ *Drake covets loons. Drake covets cash. Drake covets fame.*

Antimetabole: Repetition of words in reverse order.

(7) $[W]_a \ldots [W]_b \ldots [W]_b \ldots [W]_a$ *Drake loves loons. Loons love Drake.*

Apocope: Word-terminal ellipsis, in which the final sound or syllable of a word is omitted.

(8) $[\ldots \oslash]/[\ldots P]$ *yank / yankee*

(9) $[\ldots \oslash]/[\ldots Sy]$ *doc / doctor*

Assonance: The repetition of vowels in proximal syllables or words.

(10) $\{\ldots V_a \ldots V_a \ldots\}$ *anaphora*

(11) $\{\ldots V_a \ldots\} \ldots \{\ldots V_a \ldots\}$ *Drake covets lovely cash.*

Asyndeton: The deliberate omission of conjunctions within a series of related clauses.

(12) $\{Cl_a \ldots Cl_b \ldots Cl_c \ldots\}$*I came, I saw, I conquered.*

Consonance: The repetition of consonants in proximal syllables or words.

(13) $\{\ldots C_a \ldots C_a \ldots\}$ *kakaphobia*

(14) $\{\ldots C_a \ldots\} \ldots \{\ldots C_a \ldots\}$ *Weak Dickensian plots.*

Diazeugma: A single subject governs several verb phrases.

(15) $< [N]_a \ldots >< \oslash_a Vb_b \ldots >< \oslash_a Vb_c \ldots >< \oslash_a Vb_d \ldots > \ldots$ *Drake covets loons, loves cash, wants fame.*

Ellipsis: Omitting an expected element.

(16) $\ldots \oslash \ldots$ *Drake likes loons, Bill geese.*

Epistrophe: Ending a series of phrases or clauses with the same word or words.

(17) $< \ldots [W]_a >< \ldots [W]_a >$ *Drake likes loons. Bill likes loons.*

Epizeugma: The verb which completes a predicate occurs at either the very beginning or the very ending of its sentence.

(18) $< Vb \ldots >$ *Surfaces daily at noon, the loon.*

(19) $< \ldots Vb >$ *The loon, daily at noon, surfaces.*

Epizeuxis (palilogia): The repetition of the same word with no others between.

(20) $[W]_a[W]_a$ *Look at the loon-loon.*

Homoioteleuton: The repetition of suffixes in proximal words.

(21) $[S_a\{M\}_a] \ldots [S_b\{M\}_a]$ *Leaping, jumping, loons.*

Hypozeugma: A verb follows a series of words or phrases that it governs.

(22) $< Na, Nb, Nc, \ldots Vb >$ *Drake, Bill, Amanda, love those waterfowl.*

Mesozeugma: The verb governing multiple subjects occurs in the middle of a construction that contains them all.

(23) $< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >< \ldots Vb_a \ldots >< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >$ *Husbands and wives, sons and daughters, loved that warbling loon; sisters and brothers, uncles and aunts.*

Ploche (ploce, repetitio): The repetition of the same word in a short span of text[6].

(24) $[W]_a \ldots [W]_a$ *Loons like what loons know.*

Polyptoton: The repetition of a word, but in a different form (i.e., the repetition of a stem, with a difference in affixes)[7].

(25) $[S_a\{M_a\}] \ldots [S_a\{M_b\}]$ *Looney loons.*

(26) $[\{M_a\}S_a] \ldots [S_a\{M_b\}]$ *Nonloon loons.*

(27) $[S_a\{M_a\}] \ldots [\{M_b\}S_a]$ *Looney nonloon.*

(28) $[\{M\}S_a\{M\}] \ldots [S_a]$ *Nonlooney loon.*

(29) $[S_a\{M\}] \ldots [S_a]$ *Looney with his loon.*

and so forth.

Polysyndeton: 'Excessive' use of marked conjunction[8].

(30) $and \ldots and \ldots and \ldots$ *Drake like loons and swans and geese.*

Prozeugma: The verb in the first of a series of clauses governs the noun phrases in the remaining clauses in the series, which correspondingly contain no verbs.

(31) $< \ldots Vb_a \ldots >< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >< \ldots \oslash_a \ldots >$ *Drake likes loons, Bill swans, and Amanda geese.*

Syncope: A medial ellipsis of phones or syllables.

(32) $[\ldots \oslash \ldots]/[\ldots P \ldots]$ *ev'ry / every*

(33) $[\ldots \oslash \ldots]/[\ldots Sy \ldots]$ *ma'am / madam*

8 A SEED ONTOLOGY OF SCHEMES

From the list of schemes above (and there are many more) we can begin to organize these schemes into a "seed" ontology using the specialization (*IsA*) relation, as well as other relationships such as near-synonymy and antonymy. We begin by classifying schemes according to whether there is a pattern of addition (*InclusionScheme*), deletion (*OmissionScheme*), or permutation (*PermutationScheme*). These concepts are placed at the top level of the ontology as specializations of the generic *Scheme* concept.

We can now add to the ontology, using organizing principles based on stylistic/linguistic features like repetition, placement, and lexical governance. One level down in the ontology we specialize *InclusionScheme* into the subclass *Iteration*, which is further refined into *PlacementIteration*. We can say furthermore that *Ploche* (simple repetition) *IsA Iteration*, while in turn *Antimetabole* and *Epizeuxis* (more-complex types of repetition) are both specializations (i.e., *IsA*) *Ploche*. Subclasses of *PlacementIteration* include *Anadiplosis*, *Epistrophe*, and *Anaphora*.

The second main category of schemes, *OmissionScheme*, has a rich network of subclasses. An *Ellipsis IsA OmissionScheme* and has two main subclasses, *PlacementEllipsis* and *GoverningEllipsis*, or *Zeugma*. *PlacementEllipsis* has three specializations, *InitialEllipsis*, *MedialEllipsis*, and *TerminalEllipsis*. At the next level down in the

[6] There is some potential for confusion with polyptoton, following, given the mushiness of the notion 'same word'.

[7] The M could be null in any expression. Commonly, this shows up in cases like *My new friend is quite friendly.* But, in cases where both Ms are null, there is confusion with ploche, as in *Most of us use margarine, but he actually uses butter to butter his toast*, where the first *butter* is a noun, the second a verb.

[8] Here we've just stipulated three, assuming that greater-than-three comes along "for the ride".

ontology, *Asyndeton IsA MedialEllipsis* (at the syntactic level), *Syncope IsA MedialEllipsis* (at the sub-lexical level), and *Apocope IsA TerminalEllipsis*. *GoverningEllipsis*, the other form of *Ellipsis*, has subclasses *Prozeugma*, *Hypozeugma*, *Epizeugma*, and *Mesozeugma*.

In addition to the *IsA* relation, we can define relationships between *Scheme* concepts based on near-synonymy, part-whole (*PartOf*), and antonymy (*IsOppositeOf*). For example, *Diazeugma IsOppositeOf Zeugma*. The organization of these basic *Scheme* concepts to form our seed ontology is shown in Figure 1.

9 APPLICATION: RHETORICAL FIGURATION IN POLITICAL SPEECHES

Political rhetoric is a particularly interesting topic for investigating formal models of persuasion, e.g., [1, 15]. Our work takes a distinctive approach in piecing together the stylistic choices made across multiple levels of linguistic description through textual analysis using stylistic and rhetorical ontologies based on an integrated model of meaning. Barack Obama's inaugural address offers a particularly rich site for our computational rhetorical approach. Beginning literally from the moment of its completion, it attracted hosts of commentators. Most notably, a controversy quickly broke out between literary critic Stanley Fish and linguist Mark Liberman. Fish, in his opinion piece for the *New York Times*, noted that the oratorical leanness was, at least in part, meant to serve a literary richness. That is, *oral* eloquence was sacrificed for *written* complexity. The speech worked best as a literate artifact, "a framework on which a succession of verbal ornaments [were] hung", so that the audience was not being "invited. . . to move forward" so much as "to stop and ponder significances", as in an art gallery. An oral performance, of course, moves through time, largely independent of the hearer, whereas a written document is arrayed in space, allowing the readers to top and ponder at their leisure. The main stylistic feature Fish correlates with this stop-and-ponder literate dimension is parataxis (Gk., "arrange side by side"), the contiguous arrangement of equivalent clauses, such as Julius Caesar's "I came, I saw, I conquered." Parataxis is frequently defined in opposition to hypotaxis (Gk., "arrange under"), the syntax of subordination, and Fish argues that "the prose of Obama's inauguration is surely more paratactic than hypotactic".

Our role here is not to adjudicate between Fish and Liberman, beyond that effective oratory can work on the page *and* on the stage, and work quite distinctly in the two modes. But their exchange forms a useful backdrop to the sort of interpretation of Obama's rhetoric that our ontology supports. We can, for instance, get at the 'rhetorical austerity' issue by way of our features, polysyndeton (Gk., "many connectives") and asyndeton (Gk., "no connectives"). An instance of the former is "these men and women struggled and sacrificed and worked", with its abundance of *and*s; of the latter, "it has been the risk-takers, the doers, the makers of things", with its absence of the expected *and*. In large part, Fish is responding to the high asyndeton/polysyndeton ratio, such that there is a relatively low number of explicit connectives in Obama's inaugural address. We can see this all the more clearly if we compare another speech of Obama's, one that does have the kind of familiar eloquence that was expected on January 20, 2009. In the inaugural address, we have a ratio of 19/4 asyndeton over polysyndeton. In Obama's famous speech on winning the South Carolina Democratic presidential primary, commonly known as the *Yes We Can* speech, the ratio is far different: 6/8. The high proportion of asyndeton in *The Inaugural* goes a long way to explaining most interpreters' sense of terseness in the speech and especially Fish's stop-and-ponder opinion. The address has several paratactic passages, but the effect of these are heightened substantially by asyndeton. From these results, we can detect differences between the speeches in both register (level of formality) and ethos (variance in roles of both speaker and audience): *Yes We Can* is much less formal than *The Inaugural*, and the speaker positions himself differently towards the audience in these two speeches, an individual candidate trying to win over the electors versus the institution of the President issuing a call-to-action and responsibility.

10 CONCLUSION

We have taken the first steps towards constructing and formalizing an ontology of rhetorical figures for use in Natural Language systems such as automated stylistic annotators and text generation systems. Many very difficult problems remain to be solved. For example, because of the complexities of English spelling, we will need somewhat elaborate notions of consonants and vowels, not just simple letter repetition. Word, phrase, and morpheme are not especially simple notions, either, but working with text we should be able to leverage all kinds of cues (blanks, hyphens, punctuation generally).

Tropes provide a wealth of semantic difficulties and even many schemes provide obstacles to formalization. Formalizing isocolon (a series of similarly structured phrases), for instance, requires figuring out how to formalize—and recognize—the notion of similarly structured phrases. In principle, the problem is pretty much the same as polyptoton, in which we have the 'same' word, in different forms (*suck*, *suckingest*, *sucked*). In isocolon, we have structurally the 'same' phrase, with at least some different constituents (*I came, I saw, I conquered.*). We might also represent the parallels between tropes and schemes using formal ontological relationships: for example, synecdoche expresses a conceptual *PartOf* operation; ploche is formally *PartOf* antimetabole. Our ultimate goal is to work out a hierarchy of rhetorical figures in the form of a formal ontology, since so many figures obviously implicate others.

Our project has far-reaching implications for Natural Language Processing, as does the general turn toward persuasion in NLP (which we might call the Rhetorical Turn, borrowing a phrase from Richard Rorty in science studies[9]. As two examples, consider the long tradition of studying ethotic figures (that is, figures which support the credibility of the speaker) and of studying pathotic figures (figures which affect the emotional quotient of the discourse). The former can be very valuable for developing or locating or managing discourses in which trust is critical. The latter can be equally valuable for sentiment analysis and management. Conversely, ineffective, inappropriate, or awkward uses of figures can be detected and ameliorated. We are at the outset of a long, and we forecast, very rich collaboration between the ancient discourse technologies of rhetorical theory and the contemporary discourse technologies of Natural Language Processing.

ACKNOWLEDGEMENTS

We would like to thank the members of the Inkpot Natural Language Research Group at the University of Waterloo for many discussions and valuable input. We are grateful to Fred Kroon for his helpful assistance in preparing the figure.

[9] Herbert W. Simons. *The Rhetorical Turn.* London: Sage, 1990.

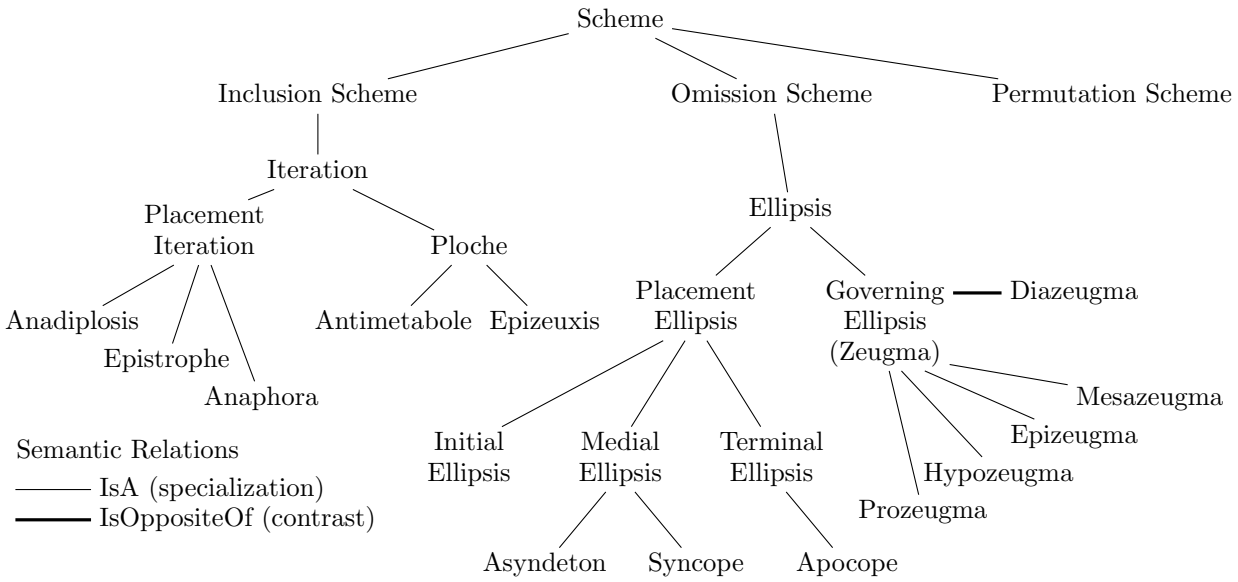

Figure 1. A seed ontology of rhetorical schemes

REFERENCES

[1] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney, 'Persuasive political argument', Proceedings of the 2005 Workshop on Computational Models of Natural Argument (CMNA V), (2005).

[2] James Crosswhite, 'Rhetoric and computation', In: C.A. Reed and T. Norman (eds.) Symposium on Argument and Computation: position papers, `http://www.csd.abdn.ac.uk/~tnorman/sac/`, (2000).

[3] Chrysanne DiMarco, Steven Banks, Olga Gladkova, and Matthew Skala, 'A fine-grained multi-level ontology for computational stylistics', Submitted to 2009 Conference of the North American Association for Computational Linguistics (NAACL), (February 2009).

[4] Chrysanne DiMarco, Peter Bray, H. Dominic Covvey, Donald D. Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Cathy Yang, 'Authoring and generation of individualized patient education materials', *Journal on Information Technology in Healthcare*, (February 2008).

[5] Chrysanne DiMarco, H. Dominic Covvey, Peter Bray, Donald Cowan, Vic DiCiccio, Eduard Hovy, Joan Lipa, and Doug Mulholland, 'The development of a natural language generation system for personalized e-health information', 12th International Health (Medical) Informatics Congress, Medinfo 2007, Brisbane, Australia, (2007).

[6] Jeanne Fahnestock, *Rhetorical figures in science*, Oxford University Press, USA, 2002.

[7] Jakub Gawryjolek, *Automated Annotation of Rhetorical Figures (working title)*, Master's thesis, University of Waterloo, expected 2009.

[8] Floriana Grasso, 'Towards a framework for rhetorical argumentation', 53–60, In: J. Bos, M.E. Foster and C. Matheson (eds.) Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG 2002), Edinburgh, UK, 4-6 September 2002, (2002).

[9] Nancy L. Green, 'Dialectical argumentation in causal domains', Proceedings of the 2008 Workshop on Computational Models of Natural Argument (CMNA VIII), (2008).

[10] Kyle Johnson, *Topics in ellipsis*, Cambridge University Press, Cambridge, 2008.

[11] Thomas S. Kane, *The Oxford guide to writing*, Oxford University Press, 1983.

[12] Fabrizio Macagno and Doug Walton, 'Argumentative reasoning patterns', Proceedings of the 2006 Workshop on Computational Models of Natural Argument (CMNA VI), (2006).

[13] William C. Mann and Sandra A. Thompson, 'Rhetorical structure theory: Toward a functional theory of text organization', *Text*, **8**(3), (1988).

[14] Chaïm Perelman and Lucie Olbrechts-Tyteca, *The new rhetoric: A treatise on argumentation*, University of Notre Dame Press, Notre Dame, Indiana, 1969.

[15] Isabella Poggi and Laura Vincze, 'Showing disinterest: A persuasive strategy to win the electors' trust', Proceedings of the 2008 Workshop on Computational Models of Natural Argument (CMNA VIII), (2008).

[16] Matthew Skala, Chrysanne DiMarco, Olga Gladkova, Elena Afros, Steven Banks, Randy Harris, and Frederick W. Kroon, 'A parser-agnostic stylistic annotator', To be submitted to *Computational Linguistics*, (February 2009).

[17] S.M. Sotillo and J. J. Wang-Gempp, 'Using corpus linguistics to investigate class, ideology, and discursive practices in online political discussions', 91–122, In: U. Connor and T.A. Upton, *Applied Corpus Linguistics*, 2004, (2004).

[18] Stephen Toulmin, *The uses of argument*, Cambridge University Press, 1958.

[19] Douglas Walton, *Informal logic: A handbook for critical argumentation*, Cambridge University Press, 1989.

Mapping Persuasive Dialogue Games onto Argumentation Structures

Andrew Ravenscroft[1] and **Simon Wells**[2] and **Musbah Sagar**[3] and **Chris Reed**[4]

Abstract. This paper reports on some prelimary research into how software tools like InterLoc can be used as an interface to the World Wide Argument Web (WWAW) and how the WWAW in return can provide a useful resource to agents acting within InterLoc. Two persuasive dialogue games, the human-oriented Critical Reasoning Game (CRG) from InterLoc and the philosophy-based agent-oriented game for permissive persuasion named PPD_0 are compared using the Dialogue Game Description Language (DGDL) as an interlingua. The expressiveness of each game is investigated by mapping output dialogues onto argumentation structures represented in the Argument Interchange Format (AIF).

1 INTRODUCTION

This paper reports on preliminary research into developing persuasive online software systems that integrate naturalistic human dialogues, thus spurring increased user engagement, with formally structured argumentation, supporting automated processing by intelligent agents and interconnection of resources online.

InterLoc [4] is software that can be used to support dialogues between groups of users and enable them to interactively explore a topic domain. This has been used thus far in a primarily educational context to facilitate debate between students. This approach can be extended through the use of intelligent tutor agents to enable new domain knowledge to be introduced into the student dialogues enabling the students to explore various paths through the topic and increasing their knowledge. Tutor agents could also be used to scaffold and direct the dialogues so that important topics were covered in sufficient depth or to ensure that the dialogue was steered towards the conclusions that the tutor wishes for the students to discover. Additionally, intelligent conversational agents could play the role of a devils advocate in an adversarial dialogue in which the students defend a given position based upon their knowledge of the domain or in which the agent tries to persuade the students to accept a position that differs from their starting position.

Recently there has also been increasing interest in online argumentation, for example MAgtALO [5] provides an interface for human-agent dialogue whereas ArgDF [3] provides an interface for constructing arguments using argumentation schemes [6]. Some aspects of online argumentation systems, for example the user facing interfaces like that in MAgtALO, suggest a good way to usefully deploy well-reasoned persuasive argumentation. By adopting natural interfaces, that support naturalistic human interaction, users can explore a problem domain, and can be supported and guided towards well-reasoned conclusions, a form of gentle persuasion technology rather than "hard sell" persusasion.

There has also been broad interest in the underlying formal representations that support the widespread sharing and interchange of argumentative resources in online systems. This is useful, not only to support the development and deployment of persuasive argumentation-based interfaces, but also to support more advanced online argumentation processing. Work towards this end has culminated in the nascent Argument Interchange Format (AIF) [1] used to record and share argument resources and a foundational element of the proposed World Wide Argument Web (WWAW) [2].

Initially our research has investigated the argumentative structures that can be extracted from InterLoc dialogues. Our aim in this task is to investigate the suitability of InterLoc as an interface to the WWAW, both as a means to elicit new arguments into the system, but also as a way for humans to explore existing WWAW argument resources using a naturalistic interface. Further to this we have investigated dialogue games from argumentation theory that are useful for implementing agent argumentation, in particular we have investigated dialogue game protocols that can be used to regulate persuasion dialogues according to the criteria of Walton and Krabbe [7].

The aim is to be able to incorporate argumentative intelligent agents into WWAW interfaces, such as InterLoc, without compromising the human friendly aspects of the current InterLoc interaction protocols. To achieve this we aim to balance the more expressive dialogue protocols which support naturalistic human dialogical interaction against the more formally structured protocols that are used in intelligent conversational agents. By doing this we propose the construction of protocols that are sufficiently expressive to allow naturalistic human interaction without introducing significant cognitive overhead but which are also sufficiently structured and formally underpinned to support support agent interaction.

2 TECHNOLOGIES

This research integrates a number of extant technologies from the domains of educational software and argumentation theory. In our preliminary work we have drawn together a range of theoretical tools which we are exploring with the aim of assembling them into a cohesive software architecture to meet the goals discussed in section 1. Our exploration of technologies has thus far been confined to evaluating two dialogue games, the Critical Reasoning Game (CRG) [4] and the Permissive Persuasion Dialogue game (PPD_0) [7]. To support the comparison and analysis of these games other technologies have been adopted, the Dialogue Game Description Language (DGDL) for describing disparate protocols using a common language, and the

[1] London Metropolitan University, email: a.ravenscroft@londonmet.ac.uk
[2] University of Dundee, email: swells@computing.dundee.ac.uk
[3] London Metropolitan University, email: m.sagar@londonmet.ac.uk
[4] University of Dundee, email: creed@computing.dundee.ac.uk

Argument Interchange Format (AIF) for representing the outcome dialogues produced from the dialogue games.

CRG is an interaction protocol used in InterLoc. This is a human-oriented dialogue game based interaction protocol which specifies a permissive, free-ranging dialogue between numerous locutors. An advantage of this protocol is that it is flexible, expressive and permissive and scaffolds dialogues by suggesting ways that the dialogue *can* develop rather than ways that it *must* develop. However this syntactical permissiveness and the lack of a semantic model provides little structure with which to support an intelligent agent interacting with the human locutors in an InterLoc dialogue.

Dialogue games used in agent communication typically incorporate commitment models. One such dialogue game developed to model permissive persuasion dialogues is PPD_0 which incorporates commitment stores which are used to track the commitments of players with respect to the locutions uttered during a dialogue. In PPD_0 the legal sequences of locutions are defined both in terms of the set of locutions that may follow an earlier locution and also the commitment state of the players. Because of these kinds of rule PPD_0 is relatively heavyweight and restrictive in comparison to CRG but this means that at any given point in a PPD_0 dialogue the set of things that can be said, and therefore the set of alternative utterances that an agent must select from, is much smaller.

Both CRG and PPD_0 are, in their original formulations in [4] and [7] respectively, specified using different mechanisms. CRG is expressed using XML and PPD_0 is expressed in a natural language description. To aid in the comparison of the games, and to reduce the complexity of transcribing rules from one game to another, the Dialogue Game Description Language (DGDL) [8] was used as an *interlingua*. DGDL is a domain specific language for describing dialogue games whose syntax is underpinned by an EBNF grammar. This supports the rapid development of syntactically correct dialogue game descriptions that can be deployed in agent software.

3 THE DIALOGUE GAME DESCRIPTION LANGUAGE (DGDL)

Communication is an important topic within intelligent agent research and is a fundamental factor in the development of robust and efficient multiagent systems. Similarly, argumentation has been recognised as a key component of an agents ability to make decisions using complex, dynamic, uncertain, and incomplete knowledge. Dialectical games are a type of multi-player argumentative dialogue game and provide a mechanism for communication which incorporates argumentative behaviours. However there have been very few tools for working with these games and little agreement over how they should best be described, shared, and reused. The Dialogue Game Description Language (DGDL) [8] is a domain specific language for describing dialectical games and provides a grammar for determining whether a game description is syntactically correct and thus provides a foundation for new tools to support the future development and wider exploitation of dialectical games.

The DGDL grammar supports the syntactically correct description of a wide array of dialectical games, whether extant games or wholly new formulations of rules. Games are described in terms of their *composition*, including specification of participants, turn structure, and commitment stores, their *rules*, regulations that manipulate the game components indirectly, and their *interactions*, the moves that players can make that directly manipulate game components.

To support the comparison of PPD_0 and CRG, the original natural language description of rules was formalised into a DGDL game description as follows:

```
PPD0{
  {turns, magnitude:multiple, ordering:strict };
  { roles, {Speaker, Listener} };
  {players, id:black, roles:{ Speaker } };
  {players, id:white, roles:{ Listener } };
  {store, id:Assertions, owner:black, structure:set, visibility:public };
  {store, id:Assertions, owner:white, structure:set, visibility:public };
  {store, id:Concessions, owner:black, structure:set, visibility:public };
  {store, id:Concessions, owner:white, structure:set, visibility:public };
  {store, id:Dark, owner:black, structure:set, visibility:private };
  {store, id:Dark, owner:white, structure:set, visibility:private };

  {Commencement, scope:initial,
    { move(mandate, next, Assertion, Speaker) } };
  {SpeakerWins, scope:turnwise,
    { if { inspect(in,{p},Assertions,Listener,initial)
    & inspect(!in,{p},Assertions,Listener,current) }
    then { status(terminate,PPD0), assign(speaker, winner) } } };
  {ListenerWins, scope:turnwise,
    { if { inspect(in,{p},Assertions,Speaker,initial)
    & inspect(!in,{p},Assertions,Speaker,current) }
    then { status(terminate,PPD0), assign(listener, winner) } } };

  {Assert, {p},
    { store(add, {p}, Assertions, Speaker)
    & store(add, {p}, Assertions, Listener) } };
  {Concede, {p},
    { if { { inspect(!in, {p}, Concessions, Speaker)
    & inspect(in, {p}, Assertions, Listener) }
    || { inspect(!in, {p}, Concessions, Speaker)
    & { event(last, Request, {p})
    || event(last, Serious, {p}) } }
    then {store(add, {p}, Concessions, Speaker) } } };
  {ElementaryArgument, {p, Q},
    { if { inspect(!in, {p}, Concessions, Listener)
    & event(past, Challenge, {p}, Listener) }
    then { store(add, {p}, Assertions, Speaker)
    & store(add, {p}, Concessions, Speaker)
    & store(add, {Q}, Assertions, Speaker)
    & store(add, {Q}, Concessions, Speaker)
    & store(add, <{p}, Q>, Assertions, Speaker)
    & store(add, <{p}, Q>, Concessions, Speaker) } } };
  {Request, {p},
    { if { inspect(!in, {p}, Concessions, Speaker) }
    then { move(mandate, next, Concede, {p})
    || move(mandate, next, WeakRetraction, {p}) } } };
  {Serious, {p},
    { if{ inspect(!in, {p}, Dark, Listener)
    & { event(last, WeakRetraction, {p})
    || event(last, Challenge, {p}) } }
    then { move(mandate, next, Concede, {p})
    || move(mandate, next, WeakRetraction, {p}) } } };
  {Resolve, {p, q},
    { if { inspect(in, {p}, Concessions, Listener)
    & inspect(in, {q}, Concessions, Listener) }
    then { move(mandate, next, WeakRetraction, {p})
    || move(mandate, next, WeakRetraction, {q}) } } };
  {Challenge, {p},
    { if { inspect(in, {p}, Assertions, Listener)
    & inspect(in, {p}, Concessions, Speaker)
    & event(!past, Challenge, {p}, Listener) }
    then { move(mandate, next, ElementaryArgument, {p, Q})
    || move(mandate, next, WeakRetraction, {p})
    || move(mandate, next, StrongRetraction, {p}) } } };
  {WeakRetraction, {p},
    { if { { inspect(!in, {p}, Dark, Speaker)
    & event(!past, Serious, {p}, Listener)
    & { event(last, Request, {p}, Listener)
    || event(last, Serious, {p}, Listener) } }
    || { inspect(!in, {p}, Dark, Speaker)
    & event(!past, Serious, {p}, Listener)
    & inspect(in, {p}, Concessions, Speaker) } }
    then { store(remove, {p}, Assertions, Speaker)
    & store(remove, {p}, Concessions, Speaker) } } };
  {StrongRetraction, {p},
    { store(remove, {p}, Assertions, Speaker) } };
  {EndTurn, {p},
    { assign(Speaker, Listener) & assign(Listener, Speaker) } };
}
```

Similarly, the XML description of the CRG rules was reformulated into a DGDL description thus:

Unfortunately the CTG DGDL description is quite long and takes up several pages due to its extensive set of locutions. For this reason, only a short but representative extract is presented here which includes the header describing the games components and a selection of interaction rules:

```
CRG{
  { turns, magnitude:single, ordering:liberal };
  { players, min:1, max:undefined };
  { player, id:$PlayerID$, role:speaker };

  {Initial, scope:initial,
    { move(,next,Suggest1) || move(propose,next,Suggest3)
    || move(propose,next,Suggest6) } };

  {Suggest1, {p}, "My idea is", { move(propose,next,Suggest3)
    || move(propose,next,Check6) || move(propose,next,Agree2)
    || move(propose,next,Transform6) || move(propose,next,Agree4) } };
  {Suggest2, {p}, "Just imagine", { move(propose,next,Suggest3)
    || move(propose,next,Check6) || move(propose,next,Agree2)
    || move(propose,next,Transform6) || move(propose,next,Agree4) } };
  {Suggest3, {p}, "What if", { move(propose,next,Agree3)
    || move(propose,next,Transform6) || move(propose,next,Agree2)
    || move(propose,next,Agree4) || move(propose,next,Check8) } };
  {Suggest4, {p}, "How about", { move(propose,next,Check6)
    || move(propose,next,Agree3) || move(propose,next,Transform6)
    || move(propose,next,Agree4) || move(propose,next,Suggest3) } };
  {Suggest5, {p}, "I feel", { move(propose,next,Suggest3)
    || move(propose,next,Check6) || move(propose,next,Agree2)
    || move(propose,next,Transform6) || move(propose,next,Suggest3) } };
  {Suggest6, {p}, "I think", { move(propose,next,Suggest3)
    || move(propose,next,Check6) || move(propose,next,Agree2)
```

```
  || move(propose,next,Transform6) } };
 {Suggest7, {p}, "Let me say more about that", { move(propose,next,Question6)
  || move(propose,next,Check6)
  || move(propose,next,Agree5) || move(propose,next,Agree1) } };
 {Suggest8, {p}, "An example", { move(propose,next,Check6)
  || move(propose,next,Check8) || move(propose,next,Transform7)
  || move(propose,next,Agree5) || move(propose,next,Agree1) } };
 {Question1, {p}, "Why?", { move(propose,next,Suggest7)
  || move(propose,next,Transform6) || move(propose,next,Suggest5) } };
 {Question2, {p}, "Can you say more on that?", { move(propose,next,Suggest7)
  || move(propose,next,Suggest5) } };
 {Question3, {p}, "Does this connect with anything for you?",
  { move(propose,next,Suggest5) || move(propose,next,Transform2)
  || move(propose,next,Maintain2) } };
 {Question4, {p}, "What do you mean when you say?", { move(propose,next,Suggest7)
  || move(propose,next,Transform6) } };
 {Question5, {p}, "Why do you think that?", { move(propose,next,Suggest6) } };
 {Question6, {p}, "Why do you feel that?", { move(propose,next,Suggest5)
  || move(propose,next,Suggest6)
  || move(propose,next,Suggest7) } };
 {Question7, {p}, "What are the possible alternatives?",
  { move(propose,next,Transform6) || move(propose,next,Check4) } };
 {Question8, {p}, "Has anyone got another idea?",
  { move(propose,next,Suggest1) || move(propose,next,Suggest4)
  || move(propose,next,Suggest3) || move(propose,next,Suggest6) } };
}
```

The complete reformulation of CRG is available however along with a description of PPD_0 from the DGDL repository[5].

4 MAPPING DIALOGUES ONTO ARGUMENTATION STRUCTURES

To adopt InterLoc as an interface to the WWAW, arguments expressed in CRG dialogues must be mapped onto AIF and ideally the process by which this occurs should be automated so that the process of taking a CRG dialogue, extracting the arguments, and transcribing them into AIF, is low cost and doesn't require transcription by experts. Similarly for PPD_0, argument structures must be extracted from the PPD_0 dialogues and recorded in AIF.

The initial approach has been to map particular dialogical sequences onto argumentation structures which are subsequently expressed as, and recorded as AIF. For example, in the sequence *Assertion–Challenge–Elementary Argument* from PPD_0, the initial assertion can be interpreted as a position taken by one player upon an issue. The Elementary Argument then provides premises in support of the Assertion. However this does not happen in isolation within a dialogue, that the player provides a conclusion and supporting argument and that they are related utterances, but arises as a result of the dialogical interaction between the players due to the intervening Challenge move. This kind of sequence can be consistently mapped onto an AIF argument structure providing a conclusion and supporting premises, licensed by an intermediate, possibly undefined, argumentation scheme. Similarly in CRG, the sequence *Suggest-Check-Suggest* can be consistently mapped onto an AIF argument.

The following four examples explore simple dialogues on the regulation of financial institutions in both PPD_0 and CRG and map the underlying arguments onto AIF which are then compared. Dialogues I and II illustrate the *Assertion–Challenge–Elementary Argument* and *Suggest-Check-Suggest* mappings discussed earlier.

Dialogue I: PPD_0
1. B: I think there should be greater regulation of financial institutions (assert)
2. W: I'm not so sure (challenge)
3. B: Let me say more; by increasing regulation, we reduce the chance of repeating recent problems (elementary argument)

Dialogue II: CRG
1. B: I think there should be greater regulation of financial institutions (suggest6)
2. W: I'm not so sure (check6)
3. B: Let me say more; by increasing regulation, we reduce the chance of repeating recent problems (suggest 7)

It is of interest that in dialogues I and II the AIF representation of the arguments, illustrated in figure 1, is identical even though the dialogues were generated according to the rules of different games. In this case the AIF enables us to easily compare the arguments expressed in the dialogues.

Dialogues III and IV again illustrate interactions from PPD_0 and CRG that exhibit surface similarity in the actual utterances of the players but yield different AIF mappings as a result of their differing underlying dialogue game protocols.

Dialogue III: CRG
1. B: I think there should be greater regulation of financial institutions (suggest6)
2. W: I'm not so sure (check6)
3. B: So what I think you are saying is that we should not regulate at all, that's crazy! (transform1)
4. W: No, that's not what I'm saying (maintain2)

Dialogue IV: PPD_0
1. B: I think there should be greater regulation of financial institutions (assert)
2. W: I'm not so sure (challenge)
3. B: Are you serious that you don't think there should be more regulation? (extractor serious?)
4. W: I am not committed to there being more regulation! (weak retraction)

In dialogues III and IV, although the locutors have expressed similar things, the commitment model of PPD_0 results in a very different AIF representation of the arguments expressed at that stage in the dialogue as shown in figure 1. In the AIF representation of dialogue III, the argument expressed in the dialogue fragment yields a similar structure to that from dialogues I and II but the conclusion is warranted by a different argumentation scheme. The players end up in conflict causing the introduction of a conflict scheme. Although the source dialogues are superficially similar, the AIF generated from dialogue IV is very different, the retraction at turn 4 causes only the I-node for the content of turn 1 to remain. It should be noted that it appears as though in dialogue IV, arguments introduced into the dialogue by player B are being disregarded because they no longer appear in the AIF representation of the argument, *at that point* during the dialogue. This is a result of the underlying commitment model of PPD_0 which demonstrates how different arguments can be produced as a result of commitment rules, altough the locutions associated with those rules are superficially similar.

The example dialogues demonstrate clear differences between the two dialogue games, CRG and PPD_0. The rules of each game are different and this can be verified through visual inspection both of the original rules and also through comparison of the reformulations into DGDL. These games also yield different dialogues, because certain chains of locutions that are legal in a more expressive game like CRG are either prohibited or not possible in a more structured game like PPD_0. However, a further complication occurs in that the particular rules of an individual game can result in similar dialogues but different underlying argumentation structures once those dialogues are analysed for their argumentative content as demonstrated by the AIF diagrams of the example dialogues.

[5] http://www.arg.computing.dundee.ac.uk/projects/a4a/dgdl/repository/

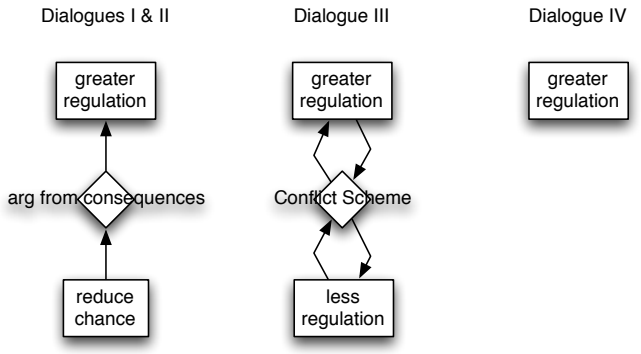

Figure 1. Fragments of AIF diagrams from the example dialogues. Dialogues I and II, produced from PPD_0 and CRG respectively, yield the same underlying AIF argumentation structure from superficially similar dialogues. Dialogues III and IV, whilst superficially similar at the dialogue level, yield different underlying AIF argumentation structures due to the effect of the commitment model of PPD_0 in dialogue IV

5 CONCLUSIONS & FURTHER WORK

It is clear that the expressiveness and naturalness of CRG dialogues contributes to user acceptance of InterLoc, an attribute that would be advantageous if InterLoc is to be used as an interface to the WWAW. However, more restrictive dialogue games like PPD_0 can prove to be better protocols for autonomous agent communication if WWAW resources are to be used to provide knowledge-bases for agents interacting within InterLoc dialogues. An ideal solution would be therefore to select and integrate elements of both games, balancing the expressiveness and naturalness of CRG against the argumentative rigour of PPD_0.

Our future work will therefore explore variant CRG games that incorporate commitment models to make the dialogues more tractable for agents whilst retaining the flexibilty of the current CRG ruleset. This will enable us to pursue the twin goals of adopting InterLoc as a WWAW interface whilst enabling InterLoc users to interact with existing WWAW resources.

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] I. Rahwan, F. Zablith, and C. Reed, 'Laying the foundations for a world wide argument web', *Artificial Intelligence*, **171**, 897–921, (2007).

[3] I. Rahwan, F. Zablith, and C. Reed, 'Towards large scale argumentation support on the semantic web', in *Proceedings of AAAI-07*, pp. 1446–1451, (2007).

[4] A. Ravenscroft, 'Promoting thinking and conceptual change with digital dialogue games', *Computer Assisted Learning*, **23**(6), 453–465, (2007).

[5] C. Reed and S. Wells, 'Using dialogical argument as an interface to complex debates', *IEEE Intelligent Systems Journal: Special Issue on Argumentation Technology*, **22**(6), 60–65, (2007).

[6] D. N. Walton, *Argumentation Schemes for Presumptive Reasoning*, Lawrence Erlbaum Associates, 1996.

[7] D. N. Walton and E. C. W. Krabbe, *Commitment in Dialogue*, SUNY series in Logic and Language, State University of New York Press, 1995.

[8] S. Wells and C. Reed, 'A domain specific language for describing diverse systems of dialogue', *Journal of Applied Logic*, **UNDER REVIEW**, (2008).

Pro or Contra? Persuasion in the Potsdam Commentary Corpus

Manfred Stede[1]

Abstract. This short paper describes our ongoing work on representing the argument structure of a particular class of persuasive texts, and on reading experiments designed to investigate the effects of certain rhetorical devices, in particular the use of explicit argumentative connectives.

1 INTRODUCTION

The first version of the Potsdam Commentary Corpus (Stede 2004) assembled 170 short texts from a German regional daily paper, which comment upon contemporary local developments. More recently, we added a set of 50 texts of a much more explicitly argumentative nature. These are taken from the Berlin-based paper *Tagesspiegel*, which every Sunday under the heading "Pro and Contra" publishes two short opinion pieces dealing with the same topic (of either local or nation-wide significance), with one arguing for and the other against the proposal, respectively.[2] Topics under discussion include whether parents should have their children vaccinated; whether Berlin should apply for the next Olympic Games; or whether some particular museum ought to be refurbished.

Consisting typically of 12-14 sentences, these pieces have to clearly make their point and provide arguments in favour of the proposition in a concise way. Nonetheless, the clear majority of these articles also mention potential counter-arguments, which are then dismissed. Thus, the argumentative structure of these articles is not overly complicated, but not trivial, either.

This short paper describes the current stage of our work with these texts: Our representation of argument structure, along with a technical annotation framework (Set. 2); a pilot study for investigating how readers perceive the persuasion in these texts (Sct. 3); and plans for experiments that study the specific effects of linguistic devices such as argumentative connectives (Sct. 4).

2 ARGUMENT REPRESENTATION AND ANNOTATION

The first phase of our work with the Pro & Contra texts aimed at identifying a framework suitable for describing the underlying arguments. We settled on the approach of Freeman (1993), who – in a nutshell – suggested a "decomposed" version of Toulmin's (1958) scheme, which allows for building representations of arguments in a piecemeal fashion. We made some minor modifications to Freeman's notation; a sample analysis of one of our texts, together with a discussion of the relationship between the structure of the argument and the linear ordering of the corresponding elements in the text, can be found in (Stede and Sauermann 2008).

The general framework of study in the Potsdam Commentary Corpus is that of *multi-level annotation*: We attach linguistic information to the texts on various levels of description (e.g., sentence syntax, coreference, rhetorical structure) in order to study the interplay of these levels and their relevance for more complicated levels of description (e.g., information structure). This approach to discourse annotation is documented in (Stede 2008). The technical infrastructure supporting it, including a standoff XML representation format as well as various software tools for performing, retrieving, visualising, and merging annotations are described in (Chiarcos et al. 2008).

Right now, we are working on the integration of the level of argument structure into this framework. We use the CmapTools[3] for drawing argument trees, and we developed a converter that maps the XML output of CmapTools to our "pivot" standoff XML format, so that the argument annotations can in turn be related to other levels of annotation in our linguistic database ANNIS. The idea is to systematically explore how the role that different portions of text play in the argumentative structure (central thesis, supporting fact, rebuttal, counter-rebuttal) are related to features of linguistic realization.

3 PILOT STUDY: WHICH TEXTS ARE CONVINCING?

As a preparatory step for experiments on the perception of persuasive elements, we conducted a pilot study to identify texts from our corpus that readers would a) understand; b) with reliable agreement identify the central thesis of the text; and c) regard as well-written and "convincingly argued".

We aimed at excluding effects of reader's pre-conceptions or prejudices and thus first selected a set of texts that cover topics which we expected our readers (Linguistics students) not to be emotionally attached to. We thus excluded texts whose topics

[1] Applied Computational Linguistics, Dept. of Linguistics, University of Potsdam, Germany. Email: stede@ling.uni-potsdam.de

[2] A similar pair of articles is published by the American paper *USA Today* under the heading "our view / opposing view".

[3] http://cmap.ihmc.us

were under public debate at the time of our study, or which are expected to be of general prominence for a student audience. At the same time, we wanted to be sure that the content of the texts was not too difficult or confusing, and hence we excluded some texts whose topic was rather idiosyncratic or relevant only to a small portion of readers of the newspaper.

Six texts meeting our criteria were given to three readers each. After subjects had read the texts, we collected them and asked the participants to fill in a short questionnaire. They were asked to judge the texts for general comprehensibility and clarity of the argument. Also, we asked the readers to recall the "central statement" of the text (which in our argument representation would correspond to the root of the tree), as well as arguments given in favour of the central statement and those against it (if any).

For two texts, the results were rather discouraging: readers gave somewhat different accounts of the central statement and also the intersection of pro- and con-arguments was low. Four texts, on the other hand, lead to good agreement. This partitioning of our six texts is now taken as the basis for our planned experiments, as described below.

4 CURRENT WORK: EXPERIMENTS ON PERSUASION

Building on the pilot study, we are currently preparing experiments aiming at finding out why the "bad" texts are being perceived in rather different ways, while the "good" texts seem to work quite well in the sense that there is a relatively uniform response by readers. Our working hypothesis is that the Freeman-type representation with its support relationships between individual statements represents the gist of the text and thereby the content that can be kept constant while changing the linguistic presentation of the argument. We thus produce variants of the texts that still adhere to the underlying argument structure but differ along two dimensions: (i) linear order and (ii) use of explicit argumentative signals. As for (i), the idea is to present arguments, rebuttals and counter-rebuttals in different orderings and check whether we can find effects in perceived persuasiveness. As for (ii), we are intrigued by a recent study by Kamalski et al. (2008), who investigated whether the use of subjective markers (especially connectives) in argumentative text has a positive or a negative effect on persuasiveness. In a nutshell, working with Dutch texts they found that presenting the argument in a more "objective" fashion tends to lead to a greater degree of persuasion. We plan to study the same phenomenon with our German texts, employing variation both in connectives (there is a difference between argumentative connectives such as *denn* and "objective-causal" ones such as *weil*; similar to English *because* versus *for*) and in modal particles and modal verbs that signal the illocutionary status of a statement – whether it is presented as a hypothesis, a fact, an appeal, and the like.

Besides, in collaboration with the psycholinguistics group in our department, we are now preparing EKP studies with shortened versions of our texts, which are also being varied along the dimensions described above. The issue is whether effects normally observed in EKP experiments for "semantically anomalous" sentences can also be found with "pragmatic surprises", e.g. an argumentative text that seems to lead to a particular conclusion, but then presents a pragmatically less compatible statement at the end of the text.

REFERENCES

[Chiarcos et al. 2008] C. Chiarcos, S. Dipper, M. Götze, A. Lüdeling, U. Leser, J. Ritz, M. Stede. A flexible framework for integrating annotations from different tools and tag sets. *Journal Traitement Automatique des Langues* (2008).

[Freeman 1991] J. B. Freeman. *Dialectics and the Microstructure of Argument*. Berlin: Foris, 1991.

[Kamalski et al. 2008] Kamalski, J. L. Lentz, T. Sanders, R.A. Zwaan (2008). The forewarning effect of coherence markers in persuasion: off-line and on-line evidence. Discourse Processes, 45, 6, 545-579.

[Stede 2004] M. Stede. The Potsdam Commentary Corpus. In: Proc. of the ACL-04 Workshop on Discourse Annotation, Barcelona (2004).

[Stede 2008] M. Stede. Disambiguating Rhetorical Structure. *Journal of Research in Language and Computation* 6:311-332 (2008).

[Toulmin 1958] S. Toulmin. *The Uses of Argument*. Cambidge University Press, 1958.

An Analysis of the Persuasive Strength of Arguments in Procedural Texts

Lionel Fontan, Patrick Saint-Dizier
IRIT, 118 route de Narbonne
31062 Toulouse Cedex France
fontan@univ-tlse2.fr , stdizier@irit.fr

1 Motivation and Aims

Argumentation (e.g. (Amgoud et al 2001, Moeschler 1985)) and, in particular, persuasive argumentation is a process frequently encountered in several types of exts where the challenge is to convince the reader to adhere to a certain point of view. Arguments come with forms of emphasis which give them more strength than normally expected, or, conversely, they may come with forms of irony or of depreciation, which influence the reader's perception of the facts associated with the arguments.

Persuation appears in different types of texts with similar objectives but with slightly different linguistic forms. This is, for example, the case in legal text analysis (Moens et alii., 2007). The situation of procedural texts, although ranging over a large set domains, seems to be simpler in terms of linguistic forms and underlying interpretation(s). One of the reasons is that procedural texts are basically action-oriented, and, therefore, the number of inferences that the user may have to do is limited as much as possible. Nevertheless, there are crucial problems associated with argumentation and persuation which are typical of procedural texts: arguments, in particular warnings, implicitly indicate that some actions are difficult to realize, and that there is a risk of failure (Dautriche et al. 2009). In terms of Action Theory, this is an interesting way to measure the complexity of a procedure and the chances to succeed, or the risks to fail.

The challenge in procedural texts is to convince the reader that the procedure which is proposed for reaching a certain goal (concrete as in do-it yourself texts, or more abstract as in social relation texts) is among the bests, that the user gets some help, hints and advices while following the procedure and that results are guaranteed, modulo some precautions (e.g. caring about warnings, reading and considering advices, carefully realizing instructions in the order they are given, etc.). It is a way of 'selling' the procedure, in comparison with other procedures describing the same task (since the web abounds in procedures, often quite different in form and contents, for realizing a certain task).

A second type of underlying objective is to make sure that the reader, when realizing the procedure, will effectively strictly and fully realize the instructions as they are given, while indicating him that otherwise he may undergo problems. In procedural texts, this is essentially realized by means of advices and warnings. It seems that these tow forms of argumentation in procedural texts follow a small number of quite standard schemas (Walton et ali., 2008). Finally, a third register in persuation, positively oriented, consists in supporting the reader when the task is complex, long or risky.

In conjunction with arguments, procedural texts abound in persuasive forms of various kinds. These forms are made visible via by a variety of marks, essentially linguistic, but also typographic, iconic or even possibly by means of images. At a global level, the presence of a number of advices and warnings in a text, is, by itself a form of persuation based on an implicit perception by the user that the text has received an in-depth elaboration and results from a long experience. Besides persuasive arguments, we observed a variety of explanation forms which have a certain implicit persuasive impact, such as reformulations, hints, definitions, etc. Besides persuation, at a theoretical level, it is of much interest to define a formal model of procedurality in terms of Action Theory (Dautriche et al. 2009). Within procedures, a number of persuasive forms also introduce some form of comfort for the user, so that he can work safely and without too much stress and worries.

2 The explanation structure in procedural texts

2.1 A global view of the explanation structure

We first constructed a quite large corpus of texts oriented towards action (about 1700 texts in French from a large number of web sites) from several domains. These texts which are, roughly, procedural texts, are quite diverse in style and complexity, from cooking, do it yourself, gardening, equipment maintenance, to social relations, health, and didactics. Those texts are in general not very long, ranging from half a page to 4 pages.

From this corpus, we established a classification of the different forms explanations may take (Fontan et al 2008). The main structures we identified are facilitation and argumentation structures. These structures are organized as follows:

- **facilitation structures**, which are rhetorical in essence (Kosseim et al. 2000, Van der Linden 1993), correspond to *How to do X ?* questions, these include two subcategories:
  (1) user help, with: hints, evaluations and encouragements and
  (2) controls on instruction realization, with two cases:
  (2.1) controls on actions: guidance, focusing, expected result and elaboration and
  (2.2) controls on user interpretations: definitions, reformulations, illustrations and also elaborations.
- **argumentation structures**, corresponding to *why do X ?* questions. These have either:
  (1) a positive orientation with the author involvement (promises)

or not (advices and justifications) or
(2) a negative orientation with the author involvement (threats) or not (warnings).

In procedural texts, we essentially observed advices and warnings since there is seldom any involvement from the author.

User help structures aim at making the user more omfortable with the current document: the way hints (*prefer a sharp knife*) and encouragements (*at this stage you've done the difficult part*) are termed and are perceived by the reader is a crucial step in the persuation process. Evaluations are in general accurate and positively oriented, guiding the user and preventing him from any questioning and discouragements (*now your sauce must look yellow, if not add more flour*). User guidance and controls on user interpretation provide the necessary assistance (possibly user parameterized, depending e.g. on how much interactions the user wishes, the type of help it requires, etc.) to guarantee a certain success, in particular when the procedure is difficult or long, with several subparts. This contributes to a feeling of control and safety w.r.t. actions being realized.

2.2 Arguments in Explanation Structures

Arguments in procedural texts serve very different purposes. They make explicit the risks that the user may undergo if he does not follow the instructions, its responsability is clearly made explicit and his role is more active. In terms of persuation, the strength of the arguments and the illocutionary force of the statements aim at convincing the reader of the reality and the importance of the risks, in the case of warnings, or of the gains in the case of advices.

It is important to note that all these aspects do not operate in isolation, but they all contribute to the success of the procedure realization. For example, well designed hints will convince the reader that the document is of high quality and that, therefore, warnings should be taken seriously. An example, using the square bracket notation, of such a structure within an instructional compound is:

[$_{instructional\ compound}$
[$_{Goal}$ To clean leather armchairs,]
[$_{argument:advice}$
[$_{instruction}$ choose specialized products dedicated to furniture,
[$_{instruction}$ and prefer them colorless]],
[$_{support}$ they will play a protection role, add beauty, and repair some small damages.]]]

We have here an argument of type advice which is composed of two instructions (or conclusions) and a conjunction of three supports which motivate these two instructions.

The explanation structure is realized by language expressions, characterized by dedicated linguistic marks typical of help statements, reformulations, etc. The typography is also an important factor via the ease of readability it introduces and also by the professionalism it suggests. Obviously, the impact to the layout in general is very difficult to measure. Our goal is to identify and categorize most of these marks, and then to a priori sort them on various scales related to persuation strength, so that, ultimately, the parameters of persuation can be measured on a given procedural text, instruction by instruction. It is also crucial to evaluate how these elements are interpreted by a variety of users. It is obviously difficult to derive a formal model due to the subjectivity of the measures (Grosz et al. 1986): in this short document, we focus on argument strength identification.

3 Processing arguments

3.1 Processing warnings

Warnings are basically organized around an 'avoid expression' combined with a proposition. The variations around the 'avoid expression' capture the illocutionary force of the argument, ordered here by increasing force, the latter expression being very strong. We give below, for the the three major classes we have observed, the basic pattern (between quotes) for the conclusion part of the argument (which has the form of an instruction), an example and the frequency observed in our corpus:

1. 'prevention verbs like avoid' (NP / to VP) (*avoid hot water*), (frequency: 48%)

2. 'do not / never / ... VP(infinitive) ...' (*never put this cloth in the sun*), (frequency: 36%)

3. 'it is essential, vital, ... to never VP(infinitive)', *it is vital to never take this medicine at the beginning of the meal*, (frequency: 6%).

Supports for warnings convey statements with a negative polarity. These are identified and delimited from various marks:

1. connectors with a negative orientation such as: *sinon, car, sous peine de, au risque de* (otherwise, under the risk of), etc. verbs expressing a consequence or verbs in the conditional form (*could damage...*),

2. negative causal expressions of the form: *in order not to, in order to avoid, etc.*

3. specific verbs such as risk verbs introducing an event (*you risk to break*). In general the embedded verb has a negative polarity.

4. very negative terms, such as: nouns: *death, disease, etc.*, adjectives, and some verbs and adverbs.

We built a lexicon of about 200 negative terms found in our corpora. While forms (1) and (2) are quite standard, those in (3) and (4) are much stronger, they appear in our corpus in about 28% of the situations. As reported in (Fontan et al. 2008), we carried out an indicative evaluation (e.g. to get improvement directions) on a corpus of 66 texts over various domains, containing 262 arguments. Those texts where manually annotated by a trained linguist, and the results were then compared with the system output. We get the following results for warnings:

conclusion recognition	support recognition	(3)	(4)
88%	91%	95%	95%

(3) conclusions well delimited (4) supports well delimited, with respect to warnings correctly identified.

3.2 Processing Advices

Conclusions of type advice are essentially identified by means of two types of patterns (English glosses given here):

1. advice or preference expressions followed by an instruction. The expressions may be a verb or a more complex expression: *it is advised to, prefer, it is better to, preferable to, etc.*,

2. expression of optionality or of preference followed by an instruction: *our suggestions: ...*, or expression of optionality within the instruction (*use preferably a sharp knife*).
3. very negative terms, such as: nouns: *death, disease, etc.*, adjectives, and some verbs and adverbs.

Supports of type advice are identified on the basis of 3 distinct types of patterns:

1. 'Goal exp + (adverb) + positively oriented term'. Goal expressions are e.g.: *in order to, for*, whereas adverb includes: *better* (in French: *mieux, plus, davantage*), and 'positively oriented term' includes: nouns (*savings, perfection, gain*, etc.), adjectives (*efficient, easy, useful*, etc.), or adverbs (*well, simply*, etc.). We constructed a lexicon of positively oriented terms that contains about 50 terms. Not surprinsingly, positive terms are far less numerous than negative terms.

2. Goal expression with a positive consequence verb (*favor, encourage, save*, etc.), or a facilitation verb (*improve, optimize, facilitate, embellish, help, contribute*, etc.),

3. the goal expression in (1) and (2) above can be replaced by the verb 'to be' in the future: *it will be easier to locate your keys*.

4. very negative terms, such as: nouns: *death, disease, etc.*, adjectives, and some verbs and adverbs.

Advices are related to optionality or preferences. The different marks above do not introduce a priori any strong difference in terms of persuation. It seems that if some terms look stronger than others, some informal experiments tend to indicate that it is more a matter of personal interpretation.

Similarly as above, we carried out an indicative evaluation on the same corpus of 66 texts containing 240 manually identified advices. We get the following results for advices:

conclusion recognition	support recognition	(3)	(4)	(5)
79%	84%	92%	91%	91%

(3) conclusions well delimited, (4) supports well delimited, both with respect to advices correctly identified. (5) support and conclusion correctly related.

A short example of an informally annotated arguement is given in Fig. 1 hereafter. We plan to use norms, as suggested in the AIF project (Chesnevar et ali. 2007).

4 Linguistic Marks of Argument Strength

Let us now review marks related to the 'illocutionary' force of an argument, contributing to its persuasive effect, in addition to the intrinsic force of arguments presented in the classifications above. These marks can be combined with the basic patterns given in the previous section. The categories given below are a priori identical for any kind of argument, positive (rewards and advices) or negative (threats or warnings). We concentrate here on those criteria that reinforce the persuasive effets, their absence could lower these effects in some cases, but this is also a matter of style.

The criteria and evaluations given below emerged from a few unformal experiments carried out on readers in our lab:

- **Number of supports**: a conclusion associated with several explicit supports seems to be stronger than if it has just one: *do not open the door when washing is ongoing*). The strength of a conclusion with no supports is quite difficult to evaluate: in a number of cases, the support is not mentioned because it is obvious for the reader and would sound odd or verbose otherwise: *do not water your plants when the temperature is below zero degrees (not mentioned: because this may 'burn' the leaves)*.
- **Supports associated with some forms of rhetorical developments**. We observed, especially in large public texts, the presence of segments of texts in a rhetorical relation with the argument support (Mann et al. 1988, Van der Linden 1993). Among the most frequently encountered relations we have: exemplification, elaboration, development and reformulation: *because you risk to break the connectors which cannot then be repaired*, with here a kind of development (but such relations may be difficult to assign unambiguously).
- **Position of supports in the argument**: a left-extraposed argument is stronger than when it appears at the end of the argument. This is a general rule in pragmatics, where left extraposed elements gets higher focus, since this position is not the expected one.
- **Typography and punctuation**: we identified several marks of emphasis: capital letters, large size, italics, bold, underlined, etc. Exclamation marks are also frequent (*do not leave in a humid place!*). However, typography and punctuation mark strength is relative to their global use in the procedure. If they appear exceptionally in an instruction, then they get more strength. In general procedures, except for video game solutions and similar types of texts, are quite sober and make a very limited use of punctuation.
- **Icons and other devices**: In a number of large public documents, extra-linguistic signs such as icons are very rich and very suggestive. There are many categories such as road signs, faces, etc. Their strength is important, but quite difficult to measure. As above, a profusion of these signs lowers their impact.
- **Marks of negation**: some marks of negation are stronger than others: 'never' is stronger than 'do not', *never use X*, *do not use X* and at the lower level we have advice verbs combined with a negation *we do not advise you to use this paint*.
- **Dedicated forms**: *pay attention:*, *important:*, *advice:*, *etc.*, these forms are close to icons. They are often highlighted.
- **Adverbs of intensity**: adverbs of intensity (e.g. *very* or of affimation (e.g. *certainly*), when applied to action verbs also introduce levels of strength *we strongly advise you not to buy...*, *this will certainly break ...*.

We also noted forms that weaken the argument. For example, the presence of a positively oriented support and a negatively oriented one for a given instruction shows the pros and cons without developing too strong a positive or negative orientation. This may be viewed also as a subtle form of persuation where a kind of objective analysis is provided to the reader.

The above linguistic marks are quite stable over a large set of types of procedural texts. Some are more frequent in some types of texts, for example, marks related to typography and text visualisation are more frequent on the web for large public audiences. Those marks can be combined to stress supports more strongly. However, we observed that, in most cases, a maximum of two of these categories may be used jointly: beyond this level supports loose their effect.

For each of these categories, we can tentatively define scales, but this is quite arbitrary and subject to errors. Research in lexical se-

$< procedure >< title >$ How to embellish your balcony $< /title >$
$< Prerequisites >$ 1 lattice, window boxes, etc.$< /prerequisites >$
....
$< instructional - compound >$ In order to train a plant to grow up a wall, select first a sunny area, clean the floor and make sure it is flat.....
$< Argument >$ $< Conclusion\ att = "Advice" >$ You should better let a 10 cm interval between the wall and the lattice. $< /Conclusion >$
$< Support\ att = "Advice" >$ This space will allow the air to move around, which is beneficial for the health of your plant. $< /Support >< /Argument >$ ... $< /instructional - compound >$ .....
..... $< /procedure >$

Figure 1. Extract of an annotated procedure

mantics, originating from (Cruse 1986) proposed some schemas for organizing along scales collections of terms which exhibit various levels of strength for a given property. However, we feel that, for each domain, these scales need to be constructed from complex and heavy psycho-linguistics experiments. We indeed noted that the relative importance of the strength of terms do depend quite heavily on the domain at stake and on the author of the text and the target audience. Obviously this is a task worth pursuing over some domains.

In a text where, in general, several arguments are found, the strength of an argument must also be evaluated w.r.t. the global strength of the others. This would be a useful contribution to Action Theory.

4.1 Perspectives

In this short paper, we presented the different forms arguments and their associated persuasive forces may take in a large variety of procedural texts. We have developed several natural language patterns to recognize conclusions and supports and related persuation marks, with quite good an accuracy. Persuation marks cover a quite large spectrum of devices, from icons, punctuation, to more semantic aspects such as verb classes, and to pragmatic aspects.

This is obviously only a first step in the analysis process, since the heart of the problem is to be able to effectively measure the persuation force associated with an argument, in isolation and in relation with the other arguments in the procedure. At the moment, we can simply, based on patterns, say if the argument has a strong positive or negative orientation. We also gave a few syntactic and morphological factors that tend to reinforce this first evaluation.

Acknowledgements This project was supported by the French ANR programme, TextCoop project.

REFERENCES

[1] Amgoud, L., Parsons, S., Maudet, N., *Arguments, Dialogue, and Negotiation*, in: 14th European Conference on Artificial Intelligence, Berlin, 2001.
[2] Chesnevar, C., et alii., it Towards an Argument Interchange Format, The Knowledge Engineering Review, 2007, Cambridge University Press.
[3] Cruse, A., lexical Semantics, Cambridge University Press, 1986.
[4] Dautriche, I. Saint-Dizier, P., *A Conceptual and Operational Model for Procedural Texts and its Use in Textual Integration*, IWCS8, Tilburg, January 2009.
[5] Delpech, E., Saint-Dizier, P., Investigating the Structure of Procedural Texts for Answering How-to Questions, LREC 2008, Marrakech.
[6] Fontan, L., Saint-Dizier, P., Analyzing the explanation structure of procedural texts: dealing with Advices and Warnings, STEP conference, Venice, August 2008.
[7] Grosz, B., Sidner, C., Attention, intention and the structure of discourse, Computational Linguistics 12(3), 1986.
[8] Kosseim, L., Lapalme, G., *Choosing Rhetorical Structures to Plan Instructional Texts*, Computational Intelligence, B. Blackwell, Boston, 2000.
[9] Mann, W., Thompson, S., *Rhetorical Structure Theory: Towards a Functional Theory of Text Organisation*, TEXT 8 (3) pp. 243-281, 1988.
[10] Moens, M-F , Boiy, E. , Mochales Palau R. , Reed, C., *Automatic Detection of Arguments in Legal Texts*, in Proceedings of the Eleventh International Conference on Artificial Intelligence and Law, ACM Press, NY, 2007.
[11] Moschler, J., *Argumentation et Conversation*, Hatier - Crédif, 1985.
[12] Talmy, L., Towards a Cognitive Semantics, vol. 1 and 2, MIT Press, 2001.
[13] Van der Linden, K., *Speaking of Actions Choosing Rhetorical Status and Grammatical Form in Instructional Text Generation* Thesis, University of Colorado, 1993.
[14] Walton, D., Reed, C., Macagno, F. (eds), *Argumentation Schemes*, Cambridge University Press, 2008.

Towards an Analysis of Argumentation Structure and the Strength of Arguments in News Editorials

Bal Krishna Bal[1] and **Patrick Saint-Dizier**[2]

1 INTRODUCTION

Editorials in general fall under persuasive texts.These types of texts intend to persuade the readers over a certain issue or topic.Hence persuasion involves the use of persuasive elements(opinion words and expressions)and facts presented in the form of arguments.A closer look at editorials reveals that they consist of an argumentation structure consisting of an opening statement(also known as the conclusion in argumentation theory[6,7,12,13,15]),which in turn is supported by other statements(known as the supports in argumentation theory)for or against the conclusion.These supports as well as the conclusion can be either facts or opinions.The underlying supports for or against a conclusion may be further developed,illustrated,justified,elaborated etc.by means of text fragments,also widely known as rhetoric relations[11].

The proposed work is aimed towards analyzing the argumentation structure and the strengths of arguments in news editorials thus determining the persuasiveness inherent in the texts.The result is a discourse analysis of opinions as stated in editorial texts producing a kind of dedicated semantic representation. Ultimately, the analyzed argumentation structure would be used to construct a synthesis of positive and negative arguments on a particular topic from one or several editorials(single or multiple sources)over a common date or a span of time.Such a synthesis can provide a relatively true view of how an event has been perceived by the public in general and is of much interest to journalists, public figures and political analysts.We also will be analyzing the change in opinions taking time as an evaluative factor for change as reported in [8,9].In order to automate the analysis as well as the synthesis construction process,we would be developing a computational model that would suggest methods and appropriate techniques.The manually annotated texts and collected editorials would serve as training data and test data respectively for validating the computation model over ideal outputs[1].

Currently, the work is in its preliminary stage,primarily focused towards analyzing the different facets of supports and rhetorical relations required for an adequate semantic-pragmatic analysis of the underlying argumentation structures in editorials.In parallel with the analysis,we are also in the process of specifying tags for annotating editorials in order to establish patterns characteristic to the different facets of support.We noted that the opinions and the argumentation structure in editorials are not so apparent and structured,which makes synthesis construction a challenging task.

[1] Madan Puraskar Pustakalaya, Lalitpur, PatanDhoka, Nepal, email: bal@mpp.org.np
[2] IRIT, 118 Narbonne 31062 Toulouse, France, email: stdizier@irit.fr

2 PERSUASIVE TEXTS AND THE UNDERLYING ARGUMENTATION STRUCTURE

In order to make the text convincing or persuasive,the common practice is to follow one or more of the following strategies [14]:

1. **Use of logical and cause effective connectives like** *however, so, and,although etc.* Example - However,if we compare the present situation with the period before 2006,we have definitely come a long way.
2. **Trying to make opinions sound like facts.** Example - The year also saw the height of anarchy,impunity and lawlessness.
3. **Use of powerful adverbs and adjectives.** Example - The Post believes that the long awaited and ever elusive CA elections will take place this year,and that the country will take the course to sustainable peace and development.
4. **Use of words like** *surely,obviously,of course, definitelyetc.* Example - If the Maoists do not run away from the elections, if the recently formed and old parties in the terai live up to the promises to allow the CA elections to take place,and if the government deals with other problems appropriately,the year 2008 will definitely herald the beginning of a new Nepal.

Source: Editorial - "Year of hope",The Kathmandu Post,December 28,2007(http://ekantipur.com/ktmpost.php)

3 SUPPORT AND RHETORICAL RELATIONS IN THE ARGUMENTATION STRUCTURE OF EDITORIALS

In our semantic and pragmatic representation of editorials,the root node is a conclusion.This conclusion bears a polarity: positive, negative or neutral.It has also a date and a source.Obviously,this polarity is either inherent through explicit linguistic marks, or needs to be deduced.In the latter case,depending on the view point of the reader,the polarity may vary.

Next,the root node is associated with one or more supports.The support relation,we define,consists of six fields that represent the facets of the support.These are represented by means of attribute-value pairs. These facets are:

- The date the support has been uttered.
- The source(name of newpaper,name of utterer if it is a reported opinion).
- The orientation of the support,namely for or against the conclusion.At the moment,we are not completely satisfied with the plain

for-against dichotomy,and hence are exploring deeper modes of classification,to better capture the notion of orientation.

- The reporting level included in the support,if any.The higher,and by-default level,is the say level,characterized by verbs like: note,notice,remark,etc.with little commitment on the part of the utterer(with attribute value commitment:low),the next level,with a stronger commitment is characterized by report verbs(inform,support,reveal,pretend,etc.,commitment:high). Finally,a number of supports showing weaker commitments,characterized by the epistemic or modal verbs such as:think,believe,assume,etc.(commitment:modal(Verb)). At this level,we do not go into further classification details since these verbs entail quite debatable modal scales.
- The conditional level,when a condition introduces the support,making it relative to another statement in the conditional part. Values for that feature are conditional:yes or no.
- The strength of the argument.This value is further divided into several fields:(1)the visible and direct strength as characterized by the polarity and the inherent strength of the terms used (direct-strength:low,average,high), (2)the relative strength, elaborated from the strength of the other supports; this allows us to take into account the personal style of the writer (relative-strength:low,average, high),and,(3)the persuasion effect that takes into account social,stylistic, contextual as well as typographical aspects,in addition to level 1.

Next talking about rhetorical relations,editorials subsume a large variety of rhetorical relations.Some are more central than others. For our purpose,we have identified the following subset of relations which really do play a crucial role:

- Exemplification:illustrates a support,while giving it a higher strength and persuation effect.
- Contrast:relates two supports A and B,where A and B are both true while partly contradicting each other.They are in general linked by connectors such as nevertheless,although,but,even if,etc.
- Discourse Frame:introduces a factual statement which indicates the environment and the scope of the conclusion(time,facts,etc.), without being a support(Last year,Nepal was declared a Federal Democratic Republic).
- Justification:where B gives reasons and explains A,this relation is stronger than the explanation relation.
- Elaboration:where B is an elaboration of A if it develops or describes a part of A.
- Paraphrase:which is just another way of saying a support or a conclusion, adding strength to it.
- Cause-effect:established a causal relation between supports.
- Result:where B results at least partly,or indirectly,from A.
- Explanation: where B is an explanation for A if it indicates the reasons for A,in a quite neutral way.
- Reinforcement:where B gives a stronger weight to A by its contents.It is stronger than an elaboration,an exemplification or an explanation.In general it contains specific marks,related e.g.to confirmation,enforcement,etc.

Below,we present an example of the argumentation structure from our corpus of editorials.

Conclusion:(<Date:2007-12-28>,<Source:KTMPOST>,
<Orientation:Positive>,<Strength:High>)
[CA elections] *will take place in 2008.*

Support:(<ID:1>,<Date:2007-12-28>,<Source:KTMPOST>,
<Orientation:Positive,Support Type:For>,<Strength:Low>)
The Post *believes that* the long awaited and ever elusive [CA elections] *will take place this year.*

Rhetorical_relation:Justification(1,2)

Support:(<ID:2>,<Date:2007-12-28>,<Source:KTMPOST>,
<Orientation:Positive,SupportType:Conditional,For>,
<Strength:Low>)
If we behave responsibly,we *will be able to hold* the **[CA elections]**.

Rhetorical_relation:Justification(1,3)

Support:(<ID:3>,<Date:2007-12-28>,<Source:KTMPOST>,
<Orientation:Positive,Support Type:Conditional,For>,<Strength:Low>)
If the Maoists do not run away from elections,
*if the recently formed and old parties of the terai live up to the promises to allow***[elections]** happen...

In the example above,the conclusion is characterized by a vector that contains id, date,source,orientation and strength.The conclusion is followed by supports and rhetorical relations.The latter establishes additional information on supports.This follows that unlike a linear model of argumentation as discussed in [10],the argumentation structure in editorials are more of a connected graph model. The description of supports is also done in the same way as in the case of the conclusion.We put the referential expression inside square brackets,which binds the supports to the event reported in the conclusion.It should be noted that **[CA elections]** and **[elections]** are the referential expressions in the example above.Similarly,the underlined text portions above are the opinion anchors,i.e.,those terms that a priori mark the statement as an opinion.For the strength, we are currently only considering the attribute, direct-strength. Other two attributes would be gradually incorporated.

As noted in the introduction section,the argumentation structure in editorials generally consists of two parts,viz.,conclusion and the supportive arguments. However,we have observed that in some of the editorials,the conclusion and the supportive arguments interchange positions.Similarly,supportive arguments are found to be linked or elaborated by rhetorical relations.For instance,in the example above,it is to be noted that there is a rhetorical relation of type - justification in between supports with ids 1 and 2 as well as supports with ids 1 and 3.

4 LINGUISTIC FRAMEWORK AND THE DISTINCTION BETWEEN FACTS AND OPINIONS

Since editorials are usually a mix of facts and opinions,there is a clear need to make a distinction between them.Opinions often express an attitude towards something. This can be a judgment,a view or a conclusion or even an opinion about opinion(s).Different approaches have been suggested to distinguish facts from opinions[3,4,5,16].Generally,facts are characteristic for the presence of certain verbs like *declare* and different tense and number forms of the verb *be* etc.Moreover,statements interpreted as facts are generally accompanied by some reliable authority providing the evidence of the claim, e.g.:

Fact: Both the two dates announced for the constituent assembly (CA) elections came and went without the vote taking place.

Reliable authority: *Election Commission for CA elections 2007.*
Fact: We have fewer people getting killed every day.
Reliable authority: *Nepal Police Department of Crime and Investigation. (December 2007)*

Opinions,on the other hand,are characterized by the evaluative expressions of various sorts such as the following [3]):
- Presence of evaluative adverbs and adjectives in sentences - *ugly* and *disgusting.*
- Expressions denoting doubt and probability - *may be*, *possibly*, *probably*, *perhaps*, *may*, *could* etc.
- Presence of epistemic expressions - *I think*, *I believe*, *I feel*, *In my opinion* etc.

It is obvious that the distinction between the two is not always straightforward.Facts could well be opinions in disguise and,in such cases,the intention of the author as well as the reliability of information needs to be verified. In order to make a finer distinction between facts and opinions and within opinions themselves,opinions are proposed for gradation as shown below:

Opinion type	**Global definition**
Hypothesis statements	Explains an observation.
Theory statements	Widely believed explanation
Assumptive statements	Improvable predictions.
Value statements	Claims based on personal beliefs.
Exaggerated statements	Intended to sway readers.
Attitude statements	Based on implied belief system.

5 STRENGTH OF ARGUMENTS

The work reported in [5]employ gradability of adjectives as a major factor for determining the strength of opinions or grades of subjective expressions in sentences.We extend this idea to arguments for which we formulate attribute-value pairs depending upon whether the argument falls under facts or opinions.Opinion arguments are further subcategorized into three groups, *High*, *Medium* and *Low* depending upon the different forms of opinion adjectives they contain.For instance,if the *Argument type* is a *Fact*, then it's strength would be *High*. Similarly,any opinion expression like *large audience* would receive the value of strength equal to *Low*, whereas (*larger audience* and *the largest audience* would receive *Medium* and *High* respectively. However,this is just a general purpose scheme and more precise specifications are necessary for dealing with more complex expressions,whose strength cannot alone be determined by the proposed technique.

To make the determination of the strength of opinions more precise and accurate, we categorize opinion words and expressions collected from our corpus into prototypically positive and negative sets as discussed in [4,16].Moreover,as part of the categorization, we propose to group semantically similar members under different subsets within the bigger sets. The smaller subsets would bear the name from one of the members belonging to that particular subset.For instance,the negative set {poor(adj), misery(n),miserable(adj),trouble(n),troublesome(n),troubling(adj)}could be named as the *misery* set.Such a subcategorization would ease in the clear cut determination of the polarity of opinion words and expressions.

Next the categorized sets would be further subject to subcategorization on the basis of the strengths as exhibited by the members of the sets.For this purpose, we further split the strength attribute into three sub-attributes,namely direct-strength,relative-strength and persuasion strength,which respectively take one of the values low, medium or high.The entries of the subsets and consequently the sets are then arranged in the same manner as in the case of operators separated by commas in a precedence table.In (Bal & Saint-Dizier, Forthcoming),we provide a detailed specification of this process.Our approach discussed above is somewhat different from [2],which takes into account the predefined values and beliefs of the readers as a crucial factor in persuasion.

6 TEXT COLLECTION AND ANNOTATION

Editorials have been collected from at least three different sources.The collected texts serve as a corpus for our research work.The editorials represent a common theme - *Soci-political* and subtheme *Peace and stability* and are taken from different dates towards the end of the year 2007 and the beginning of 2008 amounting a total of 300 plus text files,with a total of approximately 6000 sentences and an average of 20 sentences per editorial. The texts are taken respectively from *The Kathmandu Post Daily*, http://ekantipur.com/ktmpost.php, *The Nepali Times Weekly*, http://nepalitimes.com.np and *The Spotlight Weekly*, http://nepalnews.com/spotlight.php.Two annotators having a fairly good understanding of the English language have been involved in the annotation work. The annotators have been assigned the same texts to see how semantic annotations can differ among annotators.Results have shown that the difficulties in the manual annotation exist at two levels,the first one in determining the orientation of polarity of words or expressions and the second one in evaluating their strengths for the three different strength attributes - direct-strength, relative-strength and persuasion-strength.Wherever the annotators have confusions about providing one particular value,they have been advised to provide multiple values separated by commas.

For the annotation purpose, we have developed a semantic tagset, subject to further extension or modification in future. The current tagset can be represented as a list of parameters and their possible values as shown below:

Parameters	**Possible values**
argument_type	Support, Conclusion,Rhetorical_relation
expression_type	Fact, Opinion, Undefined
fact_authority	Yes,No
opinion_orientation	Positive, Negative, Neutral
orientation_support	For,Against
id	Id number of the support
date	Date of publication of the editorial
source	Source or name of the newspaper
commitment	Modal,Low,High
conditional	Yes,No
direct-strength	Low,Average,High
relative-strength	Low,Average,High
persuasion-effect	Low,Average,High
rhetoric_relation type	Exemplification,Contrast
	Discourse Frame,Justification
	Elaboration,Paraphrase
	Cause-effect,Result
	Explanation,Reinforcement

The tagset has been used to annotate the texts in XML format for outlining the argumentation structure and strength of the argument.Below,we provide a sample of the annotated text.

<Arguments>

<Conclusion date="Jan 05,2006" source="The Kathmandu Post" orientation="Negative">
The actions of the power centers seems to be heading to confrontational poll and politics.
</Conclusion>

<Support id="1" date="Jan 05,2006" source="The Kathmandu Post" orientation_support="For" commitment="Modal" conditional="No" direct-strength="High" relative-strength="High" persuasion-effect="High">
It is utter naivety on the part of the royal government to believe that the municipal polls would fix all the problems of the country.
</Support>

<Rhetoric_relation type="Elaboration" has_relation_to_support="1">
In fact,it will neither lessen the woes of the country, nor will it give any legitimacy to the autocratic monarchy.
</Rhetoric_relation>

<Support id="2" date="Jan 05,2006" source="The Kathmandu Post" orientation_support="For" commitment="High" conditional="No" direct-strength="High" relative-strength="High" persuasion-effect="High">
The polls will only widen the rift between the political parties and the royalists, and it is certain to invite more bloodshed.
</Support>

<Support id="3" date="Jan 05, 2006" source="The Kathmandu Post" orientation_support="For" commitment="High" conditional="No" direct-strength="High" relative-strength="High" persuasion-effect="High">
The government has remained obdurate,showing no intention to reciprocate the peace initiative taken by the Maoist rebels.
</Support>

<Rhetoric_relation type="Elaboration" has_relation_to_support="3">
The unilateral cease fire announced by the Maoists expires today.
</Rhetoric_relation>

<Rhetoric_relation type="Elaboration" has_relation_to_support="3">
It is unfortunate that the royal government is not persuaded by any level of persuasion to work for peace.
</Rhetoric_relation>

<Support id="4" date="Jan 06,2006" source="The Kathmandu Post" orientation_support="Against" commitment="High" conditional="No" direct-strength="High" relative-strength="High" persuasion-effect="High">
Currently the seven-party alliance is dead against the holding of the municipal polls scheduled for February 8.
</Support>

</Arguments>

ACKNOWLEDGEMENTS

I am very grateful to Professor Patrick Hall,my PhD co-supervisor,for his continuous support and guidance.I am also grateful to my institution,Madan Puraskar Pustakalaya for all the support and platform provided to me.This work was partly supported by the French Stic-Asia programme.

REFERENCES

[1] Bal B.K.,"Towards an Analysis of Opinions in News Editorials: How Positive was the year?,"in Proceedings of the 8'th International Conference in Computational Semantics,Tilburg,Netherlands,2009, pp. 260-263.

[2] Bench-Capon T.J.M.,"Agreeing to Differ:Modelling Persuasive Dialogue Between Parties Without a Consensus About Values,"Informal Logic,vol.22,no.3,pp.231-45, 2002.

[3] Dunworth K.(2008)UniEnglish reading:distinguishing facts from opinions. [Online].HYPERLINK "http://unienglish.curtin.edu.au/local/docs/RW_facts_opinions.pdf"

[4] Esuli A.and Sebastiani F.,"Determining term subjectivity and term orientation for opinion mining," in Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics,Trento,IT.,2006.

[5] Hatzivassiloglou V. and Wiebe J. M.,"Effects of adjective orientation and gradability on sentence subjectivity.," in Proceedings of the 18th Conference on Computational Linguistics - Volume 1,Saarbrcken,Germany,2000,pp.299-305.

[6] Hoffmann M.H.G.,"Logical argument mapping: a cognitive-change-based method for building common ground,"in Proceedings of the 2nd international conference on Pragmatic web,Tilburg,The Netherlands,2007,pp. 41-47.

[7] Kirschner P.A.,Shum S.J.B.,and Carr C.S.,Visualizing argumentation:software tools for collaborative and educational sense-making.:Springer-Verlag,2003.

[8] Ku L.-W.,Lee L.-Y.,and Wu T.-H.,"Major topic detection and its application to opinion summarization.," in SIGIR 2005,pp.627-628.

[9] Ku L.-W.,Liang Y.-T.,and Chen H.-H.,"Opinion extraction,summarization and tracking in news and blog Corpora.,"in AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs, AAAI Technical Report.,2006.

[10] Manfred S.and Sauermann A.,"Linearization of arguments in commentary text," in Multidisciplinary Approaches to Discourse 2008 - MAD 08, 20-23 February 2008, Lysebu, Oslo, 2008.

[11] Marcu D.,"The rhetorical parsing of natural language texts,"in Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics,Madrid,Spain,1997,pp.96-103.

[12] Moens M.-F.,Boiy E.,Palau R. M.,and Reed C.,"Automatic detection of arguments in legal texts,"in Proceedings of the 11th international conference on Artificial intelligence and law,Stanford,California,2007,pp. 225-230.

[13] Prakken H.,Reed C.,and Walton D.,"Argumentation schemes and generalisations in reasoning about evidence,"in Proceedings of the 9th international conference on Artificial intelligence and law,Scotland,United Kingdom,2003,pp.32-41.

[14] Primary Resources:English:Text Level:Persuasive Writing. [Online].HYPERLINK http://www.primaryresources.co.uk/english/englishD10.htm #persuasive

[15] Walton D.N.,Argumentation Schemes for Presumptive Reasoning.Mahwah,NJ:Lawrence Erlbaum Associates, 1996.

[16] Wilson T.,Wiebe J.,and Hoffman P.,"Recognizing contextual polarity in phrase-level sentiment analysis.,"in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver,British Columbia,Canada,2005.