Emotional and Non-Emotional Persuasion

Maria Miceli
Institute of Cognitive Sciences and Technologies
National Research Council
Viale Marx 15 – 00137 Roma
m.miceli@istc.cnr.it

Fiorella de Rosis
Department of Informatics
University of Bari
Via Orabona 4 – 70126 Bari
derosis@di.uniba.it

IsabellaPoggi
Department of Education
University Roma Tre
Via d. Castro Pretorio 20 - 00185 Roma
poggi@uniroma3.it

Abstract

A relevant issue in the domain of natural argumentation and persuasion is the interaction (synergic or conflicting) between 'rational' or 'cognitive' modes of persuasion and 'irrational' or 'emotional' ones. This work provides a model of general persuasion and emotional persuasion. We examine two basic modes for appealing to emotions, arguing that emotional persuasion does not necessarily coincide with 'irrational' persuasion, and showing how the appeal to emotions is grounded on the strict and manifold relationship between emotions and goals, which is, so to say, 'exploited' by a persuader. We describe various persuasion strategies, propose a method to formalize and represent them as oriented graphs, and show how emotional and non emotional strategies (and also emotional and non emotional components in the same strategy) may interact with, and strengthen, each other. We finally address the role of uncertainty in persuasion strategies and show how it can be represented in persuasion graphs.

1. Introduction

Decision support was one of the first subjects upon which artificial intelligence (AI) engineers tried to venture and has been, for years, a typical example of the kind of applications of AI methods. The purpose of decision support was initially identified with the suggestion of the 'best' decision to make in a given context: no argumentative or persuasive device was introduced to reinforce the addressee's intention to follow this suggestion. Those who, subsequently, tackled the problem of argumentation and persuasion (with a not always clear distinction between the two[1]) built their research on the seminal work developed by linguists, philosophers and cognitive psychologists in this area: Toulmin (1958) first of all, but also Perelman and Olbrechts-Tyteca (1969) to mention only a few. The study of these theories enlightened, to AI engineers, the limits of applying a purely logical reasoning to this domain and the need, on one side, of considering uncertainty (Zuckerman, Jinah,

[1] There is, in our view, some confusion in the literature between the notions of 'persuasion' and 'argumentation'. Although in both cases the goal of the communication process is 'to convince somebody', by *argumentation* we will mean *inducing to believe* and, by *persuasion, inducing to do*. We will assume that, in the former case, the speaker's goal is to influence an addressee's beliefs, while in the latter the goal is to influence the addressee's intention to perform some action. Since inducing to do requires acting on the system of beliefs of the addressee (Castelfranchi, 1996), there is obviously some overlapping between the two communication processes.

Mc Conachy, and George, 2001) and on the other side of introducing argumentation schemes more refined and differentiated than logical *modus ponens* (Walton, 2000). Thanks to research on multi-agent systems (Jennings et al., 1998) and also to the Workshops on 'Computational Models of Natural Argument' (see *http://www.csc.liv.ac.uk/~floriana/CMNA4.html for the last one),* the debate was thus focused on the meaning of this term and on how attempting to simulate natural argumentation requires defining new methods for representing knowledge, reasoning on it and generating natural language monologs or dialogs One of the more recent subjects of interest in this trend of research on natural argumentation concerns widening the persuasion modes from considering 'rational' or 'cognitive' arguments to appealing to *values* and *emotional* states (Sillince and Minors, 1991; Grasso et al., 2000; Guerini et al., 2003; Poggi, 2005). According to Petty and Cacioppo (1986), who are among the most frequently quoted authors in this area, persuasion processes should be based on the hypothesis that when one is motivated and able to process a persuasive message carefully, generated cognitions tend to reflect evaluative inferences about the quality of the message content, whereas when one is not sufficiently motivated or able to process carefully, 'peripheral-route' processing is likely to occur. The second kind of processing would be produced by messages that are generally called 'emotional'. Some authors (e.g., Lisetti and Gmytrasiewicz, 2002) claim, however, that this distinction between 'rational' and 'emotional' elaboration is fictitious and that the evaluative elaboration of message content may be seen as spanning a continuum ranging from the 'objective' to the 'emotional'.

Our work builds, on one side, upon research of philosophers, psychologists and psycholinguists on argumentation and persuasion and, on the other side, upon artificial intelligence research on BDI (belief-desire-intention) agents (Rao and Georgeff, 1995) and on the theory of how intentions and commitments are produced in these agents (Jennings et al., 1998). We start from a reflection on the continuum which characterizes the various persuasion modes (from the purely 'rational' to the purely 'a-rational'), to propose a formalism which unifies the various items of this continuum. In Section 2 we will, first of all, clarify what we mean by persuasion, by examining the critical dimensions of this concept. We will propose a definition of *emotional persuasion* and will discuss the role of *appealing to emotions*. We will examine two general modes for appealing to emotions by grounding them on the strict relation between emotions and goals. We will advocate that appealing to emotions does not necessarily coincide with an 'irrational' form of persuasion, and will discuss the criteria which are usually applied to distinguish 'rational' from 'irrational' persuasion. We will then describe various (emotional and non emotional) persuasion strategies, and propose (in Section 3) a method to formalise them and to represent the various sources of uncertainty they may include (in Section 4). All along the paper, we will refer to the domain of healthy eating, in which a differentiated corpus of examples may be found, ranging from advertising to messages produced by scientific agencies delegated to the promotion of a correct behaviour in this domain. In Section 5 we will discuss some related work in the domains of both argumentation and BDI models of mental attitudes, pointing to their connections with, and differences from, our approach. Finally (in Section 6) we will draw some general conclusions.

2. A model of persuasion in terms of goals and beliefs

Before providing our definition of persuasion and emotional persuasion, we need to outline some basic notions of the model in terms of which these definitions will be given. In the following, P will denote the *persuader* and R the *recipient*, that is the addressee of P's persuasive message. We will employ the male gender for P and the female gender for R. Readers should not see, in this distinction, any hypothesis about persuasion roles in the two genders, but should consider it only as a way of simplifying description of the method.

2.1. Some basic notions

a. *Goal* and *Intention*. Our notion of *goal* is very general and basic, in terms of 'regulatory state' of a system, that is, a representation which, if possible, the system tries through its actions to liken the world to (whereas the 'perceived state' is the system's representation of the world as it is) (e.g., Miller, Galanter and Pribram, 1960; Rosenblueth, Wiener and Bigelow, 1968). This regulatory state or goal is actually a complex family, including wishes, needs and intentions. In a world where resources are bounded, not any goal is chosen for being pursued, i.e. not any discrepancy between regulatory state and perceived state induces the system to try to reduce such a discrepancy (Castelfranchi, 1996; Haddadi and Sundermeyer, 1996; Bell and Huang, 1997). This choice depends on a variety of criteria, including the perceived importance of the goals, their feasibility and the amount of resources required to accomplish them.

An *intention* is a special kind of goal, which mediates the relationship between mental attitudes and behaviour (Ajzen, 1985; Fishbein and Ajzen, 1975). It is a goal endowed with the following defining properties: it is conscious; consistent with both the agents' beliefs about its possible achievement and their other intentions; chosen, i. e., implying a decision to pursue it; and planned for. So, an intention is always about some action or plan. The decision to pursue the goal implies the agent's *commitment* to it (e.g., Cohen and Levesque, 1990). However, also an intention is not necessarily pursued. If a goal is *chosen for pursuit*, and some *planning* is being done for it, this goal is already an intention – namely, what Bratman would call a 'future-directed intention', rather than an 'intentional action' (e.g., Bratman, 1987).

b. *Activated* and *generated goals.* A goal is *active* when it is included in the agent's 'goal balance' (Castelfranchi, 1990), that is, when the agent starts to assess its importance and/or feasibility through comparison with other candidate goals, in view of its possible translation into an intention. An active goal may *become* an intention if that goal is finally chosen for pursuit. An *inactive* goal of R (that is, a goal which is currently not included in her goal balance) can be *activated* by P when, in various possible ways, P makes the goal enter into R's goal balance. A *generated* goal is a new goal, i.e., a regulatory state that comes to be newly represented in an agent's mind. Goals are generated as means for pre-existing goals (Conte and Castelfranchi, 1995). The means-end relationship between a generated goal and a pre-existing one may be either internally represented (that is, planned by R) or external to R's mind. For instance, the goal to have sex is functional to reproduction, but at the psychological level R might want to have sex just for its own sake, independently of its superordinate function. Also an intention may be *generated* as a means for a pre-existing goal, on condition that this goal is *active* in R's mind.

c. *Relationships between emotions and goals*. Emotions *monitor* and *signal* goal pursuit, achievement and failure; they *generate* goals; and finally they may *translate* into goals (Miceli and Castelfranchi, 2002).

- *Emotions monitor and signal the destiny of goals.* Emotions *signal* the (possible) achievement or thwarting of goals (e.g., Frijda, 1986; Gordon, 1987): the experiences of fear, anxiety, shame, guilt, surprise, joy, and so on, all work as 'signals' of the destiny of our goals, thus accomplishing an informative function about our relationship with the environment (e.g., Lazarus, 199; Schwarz, 1990).
- *Emotions generate goals.* Once an emotion has signalled the achievement or failure of a certain goal, generally a behavioural response follows, which implies the production of some goal (of either the approach or the avoidance type). For instance, the emotion of fear signals the presence of a possible danger, and generates the goal to avoid it. In the same vein, A's envy toward B signals (to A) that A's goal of *not being less than B* has been thwarted, and generates A's goal that *B suffers some harm* (Miceli and Castelfranchi, 2002). This kind of relationship between emotions and goals is at the foundation of what we will call the *persuasive arousal of emotions*.

- *Emotions become goals*. Agents may perform (or avoid performing) an action *in order* (not) *to feel a certain emotion*: I may give you a present to feel the joy of making you happy; or do my own duty not to feel guilty. In behaviouristic terms, emotions are often (positive or negative) reinforcements, favouring either the reoccurrence or the extinction of certain behaviours. Hence the important role emotions play in learning: a given action can be performed (or avoided) not only on the grounds of the agent's expectations about its outcome and evaluations of its costs and side-effects, but also in order to feel (or not to feel) the associated emotions. This kind of relationship between emotions and goals is, as we shall see, at the foundation of what we will call *appeal to expected emotions*.

2.2 Criteria for a definition of persuasion

A variety of views of persuasion have been suggested in the relevant literature (e.g., O'Keefe, 2002). We will now try to specify our notion with regard to the following criteria.

- *Success*. Persuading somebody commonly implies *succeeding* in influencing that person. Our perspective is different: we are interested in the frame of mind of the persuader, that is on P's planning strategies, implying his theory of the recipient's mind and of the most effective (according to P) and available means for influencing R. A persuasive strategy may happen to be effective or ineffective depending on a variety of factors, including contextual or accidental causes: irrespectively of its effects, however, it remains a persuasive strategy. Thus, by persuasion we mean a persuasive intention and attempt rather than a successful persuasion.
- *P's intentional stance*. This is a crucial ingredient; although P may accidentally influence R to do something that she would not have done without his intervention, such cases are outside our notion of persuasion.
- *Intended change of R's mental state*. P may want to make R do something in many different ways: for instance by physically forcing her (say, by giving R a kicking to make her go out of a room). For a persuasive attempt to occur, however, P should want that R intends to do the required action in virtue of some change in her mental attitudes.

Persuasion and attitude change are typically viewed as strictly interrelated in the relevant literature (e.g., Beisecker and Parson, 1972; Chaiken, Wood and Eagly, 1996; Levy, Collins and Nail, 1998). Yet, this relationship is in need of further specifications.

- *Communication*. P may intentionally change R's attitudes in many different ways; for instance, by creating the physical conditions which are conducive to certain beliefs, goals, and consequent behaviours. Suppose P sets fire to a room because he wants R to get out from it: actually, P intends to change R's mental state so as to induce her to get out; however, we doubt that this should be considered a case of persuasion. By contrast, if P says to R: "You should go out: the room is burning!", this would be more likely a case of persuasive intention. However, a change in R's attitudes, even if pursued through communication, is still insufficient to define persuasion.
- *Coercion*. P may use communication to change R's attitudes in a coercive way. For instance, P may order R to do something, or may threaten he will thwart some goal of hers if she doesn't do the required action. We do not feel like 'stretching' the notion of persuasion so as to cover such cases. Therefore, we add the further requirement that, according to P's intentions, R should intend to do the required action 'freely', i.e., independent of P's exercising his power over her (Poggi, 2005). More precisely, we might say that a minimal condition for a persuasive strategy to apply is that P wants that R intends to do the required action (*at least*) *not only because P wants her to do so*. This leads us to exclude, for instance, the *ad baculum*

argument ('argument to the club or stick'; see Walton, 1996), to the extent to which it implies, either explicitly or implicitly, the exercise of power or force by P over R.

- *Manipulation*. Some authors (e.g., Burnell and Reeve, 1984) claim that persuasion should be limited to those cases in which P 'acts in good faith', that is, in R's interest, without taking advantage of her in view of some interests of his own, and without any deceptive intent. However, we see this notion of persuasion as too narrow, and prefer to talk of either manipulative or non manipulative persuasion, depending on the content of the ultimate goal of P's persuasive strategy.

We are now ready to give our basic definitions of persuasion and emotional persuasion.

2.3 Persuasion

By persuasion we mean *an agent P's (Persuader) intention to modify, through communication, an addressee R's (Recipient) beliefs or their strength, as a means for P's superordinate goal to have R freely generate, activate or increase the strength of a certain goal and, as a consequence, to produce an intention instrumental to it, and possibly to have P pursue this intention; but the minimal condition is that R has that intention.*

2.4 Emotional persuasion

By emotional persuasion we mean a *persuasive intention which appeals to R's emotions*, in either of the following ways:

- Persuasion *through arousal of emotions*: P's intention to modify R's beliefs or their strength is a means for P's superordinate goal (super-goal) to arouse an *emotion* in R, which in turn is a means for P's further super-goal to generate a goal in R, and then an intention instrumental to it. For instance P's saying to R *How disgustingly fat you are!* is meant to provoke R's shame, which should generate R's goal of not losing her face (being so disgustingly fat) and induce, as a means for this goal, her intention to go on a diet.
- Persuasion *through appeal to expected emotions*: P's intention to modify R's beliefs or their strength is a means for P's super-goal to activate, or increase the strength of, R's *goal of (not) feeling a certain emotion*, and to induce in R an intention instrumental to this goal. For instance, *If you are kind to John* (intention), *you will feel at peace with your conscience* (activated goal) or *you will not feel guilty.*

2.4.1. Why appeal to emotions?

As Aristotle already argued, persuasion relies on the interplay of three basic ingredients: the speaker's credibility and trustworthiness — especially his moral character (*ethos*); a logical and well-reasoned argument (*logos*); and the feelings of the audience (*pathos*). Aristotle's framework supports what most people suspect intuitively — that effective persuasion often appeals to both the informational and the emotional sides. Attitudes – defined as predispositions to respond in a consistently favourable or unfavourable way to a given object (Fishbein and Ajzen, 1975) – are complex constructs composed of: predispositions to certain actions; a complex of beliefs and judgments; and emotional states associated with, or aroused by, the object of the attitude. Modifying an attitude implies modifying its three components. In particular, emotional responses are characterized by a special strength and immediacy. Under certain conditions, the emotional component seems to hold a sort of 'primacy' over the informational one; for instance the possible inconsistency between affective and cognitive components

(say, feeling hostility towards somebody and at the same time having a good opinion of him) is more likely to be resolved by changes in *cognition* rather than affect (Jorgensen, 1998).

2.4.2. Emotional persuasion is not necessarily irrational

Emotional persuasion is often considered as synonymous of *irrational* persuasion (for a critical discussion of this subject, see Lisetti and Gmytrasiewicz, 2002). We do not share such a view as long as it implies that emotional persuasion should *necessarily* be irrational. Roughly speaking, rational thinking implies the *correct processing* of information. Here 'correct' does not necessarily mean 'leading to the truth', because the information available might be insufficient or false, but a kind of processing whereby conclusions are derivable from premises and the planning activity produces plausible means-ends relationships, i.e., instrumental relationships which are grounded on the evidence available (e.g., Pears, 1984). Conversely, irrational thinking goes against the evidence provided, or draws a conclusion which is not derivable from its premises.

When distinguishing rational persuasion from irrational/emotional persuasion, the following criteria or dimensions are, more or less implicitly, called into play. On the one hand, a sort of 'cold' persuasion is opposed to a 'hot' or 'warm' persuasion. 'Cold' persuasion is characterized by the provision of 'serious' information, that is, information about serious matters and goals (like health, justice, public policy) and a formal and impersonal communication style. 'Hot' persuasion typically refers to 'futile' goals, like physical appearance or attractiveness and popularity, and uses a more personalized and informal style, with abundance of qualifying adjectives. Just to provide an example of this view, while a 'cold' advertisement of a new car would appeal to 'rational' aspects as its cost or safety, a typically 'hot', that is, emotional one "might depict the car as fun, comfortable, and possibly as sexy and exciting" (Rosselli, Skelly, and Mackie, 1995, p. 165; see also Batra and Ray, 1985). On the other hand, argumentative persuasion is opposed to non argumentative persuasion, implying that the former should be viewed as rational and the latter as irrational.

In our view, both criteria ('cold' versus 'hot', and argumentative versus non argumentative) are inadequate to distinguish between rational and irrational persuasion. More importantly, they do not allow identifying 'irrational' with 'emotional':

- *Cold vs. hot*: Why should 'hot' be made equal to irrational? Consider a 'futile' goal like *being good-looking*: if P suggests plausible means in view of such a futile goal (by saying to R something like: *If you want to be more good-looking, you should take off those ugly glasses, and wear these wonderful and comfortable contact lens*) we do not see why his message should be considered an irrational form of persuasion. Of course, one might wonder if it is rational to have (and to suggest) certain kinds of goals, but such a question proposes a further question at a superordinate level (are these goals instrumental to other goals? which ones? and so on). In the same vein, the use of qualifying adjectives has in principle no irrational property: qualifying adjectives, with the evaluative implications they convey, often serve the purpose of favouring a more vivid representation of the possible consequences of (not) doing *p*. (Think for instance of the so-called 'fear appeals'.) A possible case of 'hot' persuasion is what we have called *appeal to expected emotions*, that is, P's anticipation of the emotional state R would experience if she does (or does not) perform a required action or plan (as suggested by P). For instance, *If you eat vegetables, you will be in a good mood*. The emotional state R would experience is one of the many possible and plausible consequences (either desired or undesired) of the required action (in the example, eating vegetables). In fact, there is no difference – in terms of rationality – between *If you eat vegetables, you will be in a good mood* and, say, *If you eat vegetables, you will be healthy*.

- *Argumentative vs. non argumentative*: The argumentative/non argumentative opposition might allow a distinction between rational and *a-rational* (rather than irrational) persuasion. In fact, as already remarked, 'irrational' typically means 'contrary to the dictates of reason', which implies drawing conclusions underivable from their premises, or believing something against the available evidence. By contrast, 'a-rational' might just refer to processes which are extraneous to reasoning, and independent of its rules. Non argumentative persuasion is in our view a-rational, and also emotional. This is precisely what we call *persuasion through arousal of emotions*, which is a-rational (or, more precisely, it contains an a-rational component) as long as the emotion aroused (say, shame) *directly produces a certain goal* (say, to save one's own face) independently of any reasoning. (See below the difference between emotional and cognitive activation of goals.) Although non argumentative persuasion is a form of emotional persuasion, the latter does not coincide with the former. In fact, *emotional persuasion can be argumentative*: this is precisely the case of *persuasion through appeal to expected emotions*, where, as already pointed out, typical rules of reasoning about means-ends relationships are applied, with the sole specification that the 'ends' considered concern a special class of goals: the goal *to feel (or not to feel) certain emotions*.

To sum up, emotional persuasion cannot be distinguished from non emotional persuasion by resorting either to the rational/irrational dimension (in that the identification of 'hot' with 'irrational' is unwarranted) or to the argumentative/non argumentative distinction (in that emotional persuasion can be either argumentative or not).

2.4.3. Goal activation versus goal generation

But, if emotional is not a synonym of irrational, what is the difference between inducing goals (and then persuading) through mere beliefs vs. through aroused emotions? Beliefs cannot generate goals by themselves alone. A belief can only *activate* a pre-existing goal. The latter, in interaction with the belief, can *generate* a sub-goal. Suppose I learn that tomorrow there will be shortage of water. This belief will activate my pre-existing goal *to have water*, which will generate my goal *to stock up on water* as a means for it. The cognitive activation of goals is in fact strictly related to the typical planning and reasoning procedures about goals, means, and enabling conditions. By contrast, if a belief arouses an emotion, the latter can directly *generate* a goal, independent of any planning and reasoning, i.e., independent of any *represented* means-end relation between the generated goal and some other pre-existing goal. Suppose that the belief that *John is more intelligent than I am* arouses my envy towards John. This emotion is able to generate by itself the goal that *John suffers some harm*. True, such a goal is in fact *functional* to my goal of *not being less than John*, but this means-end relation is not (necessarily) represented in my mind, and is not the reason why I want that John suffers some harm. In fact, I may want this for its own sake (because of my envy toward John), not as a means for not being less intelligent than John. Thus, unlike the purely cognitive activation, the emotional triggering of goals is a form of direct generation of goals.

The notion that the means-ends connections of goals generated by emotions are not necessarily and explicitly represented in a person's mind is worth considering in some detail, in that it implies a particular perspective on psychological mechanisms and processes: the functional and evolutionary one, which is typical of evolutionary psychology. Evolutionary psychology tries to explain the psychological mechanisms that evolved to solve adaptive problems – such as escaping dangers and predators, finding food, shelter and protection, finding mates, being accepted and appreciated among one's conspecifics – and thus surviving and delivering one's genes to one's own offspring. From the perspective of biological evolution, emotions generate goals our ancestors had to pursue in order to answer such recurrent ecological demands (e.g., Tooby and Cosmides, 1990). And, of course, the instrumental relation between such emotion-generated goals and their functions was far from being

represented in our forefathers' minds. As an example, consider an emotion like jealousy. Jealousy is activated in both sexes by an assumption of threat to a valued relationship. But, interestingly enough, the events perceived as threatening are likely to differ in the two sexes: men's jealousy is typically activated by (beliefs about) the partner's sexual infidelity, whereas women's jealousy typically focuses on the partner's loss of emotional attachment, and the possibility that he forms an attachment to someone else (e.g., Buss et al., 1992). Evolutionary psychologists (e.g., Daly, Wilson, and Weghorst, 1982; Symons, 1979) have been able to predict such differences on the grounds of the different adaptive problems faced by the two sexes: uncertainty about genetic parenthood for men, versus loss of commitment and investment (in terms of protection and resources) for women (and their children). Thus, men's jealousy is likely to motivate goals designed to reduce the threat of not investing in their own offspring, and women's jealousy motivates goals designed to reduce the threat of being deprived (together with their children) of the partner's protection and resources. However, such goals are still unlikely to be represented as means for their superordinate functions. Even when acknowledged, the superordinate functions seldom are the *reasons* why the emotion-generated goals are pursued.
We will now tackle the problem of how the concepts defined above may be formalised, as a prerequisite for any persuasion simulation attempt.

3. Formalization of Persuasion Strategies

Let p be a variable denoting an action or plan (e.g.: p = *to eat vegetables*) and CanDo(R, p), Do(R, p) be formulae denoting (respectively) that R is able to perform p and that R performs p. Let q_1, q_2,… ,q_i,… and w be formulae denoting states of the world that may include agents such as R or P (e.g. q_5 = *R is in good health*; q_4 = *R has a high cholesterol level*,…) and e_1, e_2,… ,e_j, … be formulae denoting, in particular, an emotional state of R (e.g. e_2 = *R is in good mood*). Let us denote, with $\Diamond q_i$ states of the world, and with $\Diamond e_j$, emotional states of R which will hold in a more or less near future.

We introduce the modal operators Bel, Int, A-Goal, V-Goal and Feel, to denote the various *aspects of the mental state of agent R* which are relevant in persuasion processes, that is (respectively) beliefs, intentions, active-goals, valued-goals and feelings. The first term of these operators is an agent name (R), the second is a formula. In particular: the second term of Int denotes R performing p: (Int R Do(R, q)); the second term of Feel denotes an emotional state e_j of R: (Feel R e_j); the second term of A-Goal and V-Goal denotes a state of R (either emotional or not): (A-Goal R q_i), (V-Goal R q_i), (A-Goal R e_j), (V-Goal R e_j); this means that the goal of agent R may be either to achieve a domain state q_i or an emotional state e_j of self. The second term of Bel may be any combination of formulae denoting (present or future) states of the world and features of R, such as her abilities or action performance, with the $\wedge$, $\vee$, $\neg$, $\rightarrow$ connectives. For example: (Bel R q_i) for *R believes that* q_i; (Bel R CanDo(R, p)) for *R believes that she can perform action* p; (Bel R (Do(R, p) $\rightarrow \Diamond q_i$))) for *R believes that, if she performs action* p*, state* q_i *will hold in a more or less near future;* (Bel R ($q_i \rightarrow \Diamond q_h$)) for *R believes that, if state* q_i *holds, also state* q_h *will hold in a more or less near future;* (Bel R ($q_i \rightarrow \Diamond e_j$)) for *R believes that, if state* q_i *holds, the emotional state* e_j *of self will hold as well.*

The (Bel R CanDo(R, p)) condition synthesizes a variety of enabling conditions (Castelfranchi, 1990):

- *Power from external conditions*: for example, supposing that p is *to follow a vegetarian diet*, an external enabling condition for doing p would be living in a place where someone sells or cultivates vegetables. If R believes she cannot find vegetables, she wouldn't believe she can do p.
- *Internal power* (or power from internal conditions), which in turn can be distinguished into:

  - *Power from internal capacities*: for example, a cooking ability for eating something that should be cooked;

- *Power from choice*: one 'can' perform p only if s/he believes that doing p has a positive cost-benefit balance (that is, it does not imply renouncing some other goal which is (subjectively) more important than the goal q_i for which p is a means). For example, *eating vegetables* (p), performed in order to *lose weight* (q_i), may imply renouncing the goal to have a 'fast-made' food, in that p may imply having to wash and cook the vegetables, which one may find very boring and time-consuming.

The basic ingredients of reasoning by any persuading agent P include: (i) P's second-order beliefs about R's beliefs, value, activity and state of achievement of goals, intentions and possible actions; and (ii) first-order beliefs about the opportunity, for P, to achieve a given own goal. We assume that P believes that: (i) *if R intends to perform* p*, R will do it* (strong assumption!) and that (ii) *if R performs* p*, P will achieve his own goal* w. P's plan is therefore aimed at inducing, in R, the intention to perform p. P may apply various strategies to get this: he may *generate intentions* or *activate goals*. In the first case, a goal q_i of R is assumed, by P, to be already active, and P tries to show the reasons why R should intend p as a means for q_i. In the second case P, assuming that R's goal q_i is *not* active, tries to activate it so as to satisfy this basic condition for allowing consideration of a means-end relation between p and q_i. In both cases, P may evoke either 'rational' or 'emotional' factors, or an appropriate mixture of them.

A comment about the relationship between q_i and w is needed. As we said, p is instrumental to a goal of R (q_i) but it may be meant (by P) to favour, at the same time, the achievement of a goal (w) of himself. This relationship might represent some kind of *manipulation*, that is of P's unfair use of the persuasive message in order to achieve his own goals: in this case, to effect his purpose, P will have to conceal this relationship from R, by making her believe that q_i corresponds, as well, to his own final goal, and that his suggestion that R performs p in order to achieve q_i is 'in her interest' (see Example 1 below). However, there are also many cases in which this coincidence of interests (q_i = w) is real, and P is really acting in favour of R's interests**.**

3.1. Generating intentions by acting on 'non emotional' goals

To induce intention about p in R, P believes that the following conditions should hold: a goal q_i should exist with a sufficiently high value to R; this goal should be active; R should believe that performing p implies achieving q_i and that both internal and external conditions hold to perform it. This hypothesis may be formalised in terms of a conditional rule, which defines the relations among the components of R's mental state that should be verified, for an intentional state to hold [2]:

$$[(\text{V-Goal R } q_i) \wedge (\text{A-Goal R } q_i) \wedge (\text{Bel R } (\text{Do}(R,p) \rightarrow \Diamond q_i)) \wedge (\text{Bel R CanDo}(R,p))] \rightarrow ? \ (\text{Int R Do}(R,p)) \qquad (1)$$

The implication (1) is represented as an oriented graph in Figure 1a[3]. If combined with the relationship between R performing p and P achieving w, this graph may be employed for two kinds of reasoning:

- to simulate the reasoning process that P follows to verify whether a supposed cognitive state of R (her values, goals, abilities and beliefs about means-goals relations) will presumably enable him to achieve his goal w;
- to simulate the planning process that P performs to achieve his goal w. In this case, the subgoals of his persuasion strategy will be (i) to increase the value that R attaches to q_i, (ii) to activate goal q_i in R; (iii) to

[2] The symbol →? denotes an uncertain implication : in Section 4 we will discuss how to represent and treat it.

[3] For the sake of simplicity, we omit the diamond ◊ in all the figures.

convince R that a relation exists between performing p and achieving q_i and, finally, (iv) to convince her that internal and external conditions hold for her to perform p.

Figure 1 about here

Example 1

Figure 1 represents an example in the domain of healthy eating. In this case: q_i = *R loses weight*; p = *R eats vegetables*. Persuasive message: *You should eat more vegetables: eating vegetables helps to lose weight.* Notice that, in this example, P's goal w may either coincide with q_i or not. In the latter case, suppose w = *P avoids buying meat*, in a context like the following: P and R live together, P is the one who generally does the shopping for both; there is not meat in the fridge and P wants to avoid buying it.

3.2.Generating intentions by acting on the goal (not) to feel a certain emotion

Actions may be performed (or avoided) also in order to feel (or to avoid feeling) a given emotion. What we called (in Section 2.4) *appeal to expected emotions* is a persuasion strategy which exploits this opportunity. In this strategy, an intention may be generated by the goal of feeling (or avoid feeling) an emotion which is associated with it. For example, as we said in Section 2.1, one may give a present in order to feel the joy produced by the satisfaction or the gratitude of the receiver; one may avoid offending someone in order to avoid feeling guilty, etc. Emotional generation of intentions may be represented as follows:

$$[(\text{V-Goal R } e_j)\wedge(\text{A-Goal R } e_j)\wedge(\text{Bel R (Do(R,p)}\rightarrow\Diamond e_j))\wedge(\text{Bel R CanDo(R,p)})] \rightarrow ?\ (\text{Int R Do(R,p)}) \qquad (2)$$

Example 2

If, for instance, e_j = *R is in good mood*, the following message might be produced in a context like the one described in Example 1: *You should eat more vegetables: eating vegetables induces good mood!;* or *Do you remember? You really feel at ease when you eat vegetables.*

A second persuasion scheme (Figure 1b) may then be added to the scheme in Figure 1a. In this case, the goal of R is to be in emotional state e_j. Expected emotions may be positive, as in the previous example, but also negative: in this case, the first three conditions in (2) are substituted with the following ones:

$$[(\text{V-Goal R } \neg e_j)\wedge(\text{A-Goal R } \neg e_j)\wedge(\text{Bel R (Do(R,p)}\rightarrow \Diamond\neg e_j))\wedge(\text{Bel R CanDo(R,p)})] \rightarrow ?\ (\text{Int R Do(R,p)}) \qquad (3)$$

where e_j is a 'negative' emotional state, as in Example 3 below.

Example 3

Suppose e_j coincides with *R feels aggressive*, which is an emotional state R has the valued and active goal to avoid. The following message might be produced by P in order to induce her to intend to eat vegetables: *If you eat vegetables, you will not feel so aggressive.*

3.3. Goal activation

The messages in Examples 1, 2 and 3 are based on the assumption that R's goal - q_i or e_j - is of value and is already active in her mind. Let us now consider what happens if P assumes that this goal, though being of value

to R, is not active in her mind when P plans his persuasive message: what P will additionally have to do is to 'make something' to activate it. We will now discuss how this can be done.

3.3.1. Cognitive activation of goals

As we said in Section 2.4.3, a belief q_k (say, the belief that *tomorrow there will be shortage of water*) cannot by itself *generate* a goal q_i (say, *to stock up on water*); it can only *activate* a goal q_h (*to have water*) which is already represented in the subject's mind. Once R's goal q_h has been activated, if R believes that goal q_i is *useful* to achieve q_h, this will generate, in turn, q_i as a sub-goal. Being generated by an active goal, also q_i will be active. Cognitive activation of goals may be represented as follows:

$$(\text{Bel R } q_k) \rightarrow ? \ (\text{A-Goal R } q_h)) \quad (4)$$

$$[(\text{A-Goal R } q_h) \wedge (\text{Bel R } (q_i \rightarrow \Diamond q_h))] \rightarrow ? \ (\text{A-Goal R } q_i) \quad (5)$$

According to this model, cognitive activation is closely related with planning and reasoning on ends-means-conditions, and any goal q may be activated with this mechanism (see Figure 2a).

Figure 2 about here

Example 4

Suppose that (according to P's model of R's mind) R's belief that her cholesterol level is high is likely to activate R's goal to be in good health (q_h). Also suppose that R believes that losing weight (q_i) is a means for being in good health (q_h). P may give R the 'activating' information about her cholesterol level (q_k) so as to induce her to intend to lose weight (q_i). This may result in the following persuasive message: *Your cholesterol level is high! Maybe you are overweight.*

The activated goal q_h may be either a non emotional goal like *to be in good health* as in the previous Example, or the goal of feeling an emotion (say, *to be cheerful*), which would result in a different persuasive message, like: *How cheerful you were, when you lost weight! And now…*

3.3.2. Emotional activation of goals

As already pointed out (Section 2.1), emotions signal the (possible) achievement or frustration of goals. They also generate other goals which are *functionally* instrumental to increasing the probability of achieving the monitored goals or avoiding their thwarting, or to limiting the damages implied. For instance, envy towards another person signals that the goal of *not being less than this person* has been thwarted, and generates the goal that *this person suffers some harm*, which is functional to reducing his or her power.

Unlike what happens in cognitive activation, goals which are triggered by emotions are in fact generated without the mediation of other goals: a certain belief arouses a given emotion, and the emotion in turn generates a goal *which is not represented as a means* for some other pre-existing goal. The generated goal, being produced by an aroused emotion, is also *active* in the subject's mind. We may call this an *a-rational* process, as no planning link is *represented* between the activating conditions (the monitored goal) and the generated goal. For instance, as we said, in the mind of the subject who feels envy, obtaining another person's harm is not represented as a means for having 'not less power than' this person. Emotional (or a-rational) goal triggering is in fact a form of goal generation, and may be formalized as follows (see Figure 2b):

(Bel R q_j) →? (Feel R e_j) (6)

(Feel R e_j) →? (A-Goal R q_h) (7)

[(A-Goal R q_h) ∧ (Bel R (q_i → ◊ q_h))] →? (A-Goal R q_i) (8)

Example 5

Suppose that (according to P's model of R's mind) R's belief that she is disgustingly fat (q_j) is likely to arouse in R the emotional state to be ashamed (e_j), which in turn is likely to generate and activate the goal to save her face (q_h). Such a goal, together with the belief that being in shape is a means for saving her own face, is likely to generate R's goal of being in shape (q_i.) Here, with a message like *You are disgustingly fat*, P may trigger the whole process.

It is worth specifying that q_h, once generated, is likely to become the end-goal of a planning process. In other words, although 'a-rationally' produced in R's mind, q_h may induce a very 'rational' planning, instrumental to its pursuit and achievement. For instance, once envy has generated the goal of obtaining the other's harm, this goal may generate other goals, such as *discrediting the other* or *attacking the other* which are *represented as means* for it. In the same vein, in Example 5, once shame has generated the goal *to save face*, the latter may induce the rational planning *to be in shape*.

3.3.3. Context-dependent activation of goals

The effects of a persuader's message on the recipient depend on the context in which the message is delivered, including the specific recipient of the message (Poggi and Pelachaud, 2000).

a. A communicative act may produce, in different addressees, either a cognitive or an emotional activation of a goal q. For instance, the sentence *You look sick* can either non emotionally activate the goal to lose weight, or generate it through the emotion of fear. The latter will be more likely in a subject (say, a hypochondriac) who is very sensitive to any possible threat to her health.

b. In addition, a given belief may arouse, in different contexts, different emotions, and every emotion may, in turn, generate and activate different goals. For example, saying *You are disgustingly fat* may arouse shame, anger or a mixture of the two emotions. Saying *Maria is much more in shape than you are* may arouse envy, emulation or various mixtures of the two emotions, again according to the context. Shame may generate and activate the goal of *performing better* or else the goal of *giving up the task* (Carofiglio et al, 2005).

3.4. Combining strategies

As shown in the previous sections, even emotional persuasion is partially based on the recipient's planning ability, while, on the part of the persuader, very 'rational' strategies may appeal to the recipient's emotions. So, the distinction between emotional and non emotional persuasion is, in real life, very blurred. In addition, the strategies we represented in the subnets in Figures 1 and 2 are not necessarily alternative: if needed, they may be combined to strengthen the persuasive effect, as we will show in one of the simulation examples in the next section.

Example 6

We may combine cognitive with emotional goal activation (Example 4 with Example 5: see Figure 3) by setting: q_1 = *R loses weight;* q_4 = *R has a high cholesterol level;* q_5 = *R is in good health;* q_3 = *R looks disgusting;* e_1 = *R is ashamed;* q_2 = *R saves face.* Message: *Your cholesterol is high, and you look disgustingly fat!*

Example 7

Persuasion strategies may combine acting on non emotional goals with acting on the goal to feel emotions (see, again, Figure 3, which combines Examples 1 and 2):

p = *to eat vegetables;* q_1 = *R loses weight;* e_2 = *R is in good mood.* Message: *You should eat more vegetables: eating vegetables helps to lose weight and induces good mood!*

The strength of a combined strategy depends on the hypotheses the persuader can make about the recipient's personality. R's personality may influence the value she assigns to goals (Oatley and Johnson-Laird, 1987; Poggi and Pelachaud, 2000; Ortony, 2003), her 'propensity to feel emotions' and how she believes in her own capacities. In the next section, we will show some examples of how we represent the effect of personality traits of R in terms of the uncertainty associated with nodes and their strength.

Figure 3 about here

---

4. Role of Uncertainty

Uncertainty plays a substantial role in affecting the success of a persuasion strategy (Walton, 1990): the persuader tries to exploit uncertainty factors when planning a prospectively successful strategy in a given context. In Toulmin's argumentation schemes, uncertainty is expressed verbally in the *qualifier*, which represents the argument strength: to be persuasive, a speaker can utilize the qualifier to indicate the worth of his argument. In the famous example: *A man born in Bermuda will presumably be a British subject* (Toulmin, 1958), the word *presumably* represents the existence of a medium-strength link between the premise (to be born in Bermuda) and the conclusion (to be a British subject). Other sources of uncertainty in argumentation and persuasion originate from data, which may be reported or observed with some degree of uncertainty or may be, in turn, the result of an argumentation process (Kienpointer, 1992). Some of the *critical questions* that Walton and Reed associate with their argumentation schemes (Walton and Reed, 2002) are characterized by the uncertainty in the data or in the link between data and claim. For instance:

a in the *argument from cause to effect*: "is the *evidence* cited *strong enough* to warrant the generalization as stated?";

b in the *argument from sign*: "what is the *strength of the correlation* between A (the sign) and B (the event)?";

c in the *argument from consequences*: "*how strong* is *the likelihood* that these cited consequences will (may, must etc) occur?" (Walton, 1991).

There are, therefore, at least two features of uncertainty that should be represented in persuasion graphs: the *strength of a piece of evidence* (as in a.) and the *strength of the link* (or of the 'correlation') between premises and conclusion (Toulmin's qualifier), as in b. or c.

In our model, the persuader P applies his persuasion strategies to an image of the recipient's mental state, which is necessarily uncertain. He will then believe that R is more or less *likely* to hold a given belief or goal; he will presume the *value* R likely attaches to her goals (either rational or emotional), her likely *propensity to feel*

specific emotions and the *intensity* of the emotions probably felt by R. For these reasons, rather than representing P's reasoning on R's mind in a logical framework, we adopt a representation and reasoning formalism in which uncertainty is included and handled according to probability theory: the formalism of belief networks.

Belief networks (BNs) are a well-known formalism to represent probabilistic reasoning in directed acyclic graphs (Pearl, 1988). Random (binary or multivalued) variables are associated with their nodes and oriented arcs represent causal links from parent nodes to their children: they may represent, as well, any kind of relationship among variables. The structure of the graph implicitly defines conditional independence assumptions among the variables. A probability distribution is assigned to the variables associated with the root nodes and a conditional probability table to intermediate and leaf nodes in the network. These parameters assign a precise weight to the uncertain implication '→?', for every combination of values of the parents-child nodes. In our belief networks, some of the components of R's mental state (belief nodes, intention-nodes and goal-activation nodes) are represented as boolean variables while others (goal-value nodes and emotion-feeling nodes) are represented with multivalued variables. The model is employed to simulate how to plan a persuasion message which is suited to the particular context in which it should be delivered. The problem may be formulated as follows. Given:

- a general persuasion model which has been instantiated into the particular application domain considered,
- a plan p which P wants to be intended by R,
- a set of beliefs-means-goal relations about p, and
- a set of hypotheses about the mental state of R: her beliefs, intentions and goals, integrated with affective elements resulting from the specific context and her personality,

verify the effect produced on R by alternative persuasion strategies by means of a *what-if* type of reasoning, and select the 'strongest' strategy. The strength of a strategy is measured in terms of its *impact* on R's intentional state, that is on the probability of the instantiated node (Int R Do(R,p)). Let us illustrate this method with an example in the healthy eating domain.

a. Instantiating strategies into a specific situation

As a first step, P *instantiates* the general persuasion knowledge described in Section 3 into the selected application domain. This is done by assigning a value to every variable and sentence and by combining the strategies in a unique persuasion graph.

Example 8

Let us look at Figure 3 again. Here, the node ActOnNonEmotGoals represents a persuasion strategy which is based on acting on non emotional goals, while the node ActOnGoalFeelEmotion represents a strategy which is based on acting on the goal of (not) feeling a certain emotion. The two nodes CognitiveGoalActivation and EmotionalGoalActivation represent (respectively) cognitive and emotional activation of goals.

We represent variable and sentence instantiations in the rectangles aside the nodes. Variables and sentences are instantiated as follows: p = *R eat**s** vegetables* (EatVegs); q_1 = *R loses weight* (LoseWeight); q_3 = *R is disgustingly fat* (IsDisgFat); e_1 = *R is ashamed* (IsAshamed); q_2 = *R saves face* (SaveFace); q_4 = *R has a high cholesterol level* (HasHighChol); q_5 = *R is in good health* (IsInGHealth); e_2 = *R is in good mood* (IsInGMood).

The relationship between (Bel R q_i) and (Feel R e_j) (see (6) in Section 3.3.2) is simulated with *a model of emotion activation* (Carofiglio and de Rosis, 2003) external to the belief network: this module simulates how a given message of P (and consequently a belief of R) can activate, in R, a mixture of emotions, each with its intensity

(Carofiglio, de Rosis and Grassano, 2005). The relationship between (Feel R e_j) and (A-Goal R q_h) in (7) is simulated with a model of *emotional goal activation* which represents, again, the possible relationship between the two items.

b. Introducing uncertainty in the graphs

As we said, we quantify uncertainty in terms of *probability* and treat and combine it according to probability theory, with one of the algorithms developed for propagating uncertainty in belief networks (Spiegelhalter, 1986). Parameters are introduced so as to represent relationships of various types: for instance, two strategies may be seen as alternative or they may combine so as to strengthen each other.[4] Every node (every premise in an argumentation step) may influence the truth value of its child-node (the conclusion) with its own strength; strengths are represented, as we said, as conditional probabilities.

Example 9

If activation of goal q_1 (*lose weight*) is presumed to be influenced, for R, by emotional appeal more than by cognitive appeal, then we may set:

Prob((A-Goal R LoseWeight)| EmotionalGoalActivation $\wedge$ $\neg$CognitiveGoalActivation) = .65

Prob((A-Goal R LoseWeight)| $\neg$EmotionalGoalActivation $\wedge$ CognitiveGoalActivation) = .5

while, if activation of the goal by a combination of the two strategies (as in Example 7) may be presumed to produce a stronger effect, then we may set:

Prob((A-Goal R LoseWeight)| EmotionalGoalActivation $\wedge$ CognitiveGoalActivation) = .8

At the same time, if P presumes that cognitive activation of the goal *to lose weight* requires that R believes that LoseWeight enables IsInGHealth, then we may set:

Prob(CognitiveGoalActivation|(A-Goal R IsInGHealth) $\wedge$ Bel R (LoseWeight $\rightarrow$ $\Diamond$ IsInGHealth) = .7

Prob(CognitiveGoalActivation|(A-Goal R IsInGHealth) $\wedge$ $\neg$Bel R (LoseWeight $\rightarrow$ $\Diamond$ IsInGHealth) = .2

… and so on

c. Using the model to find out an 'appropriate' strategy

To test alternative persuasion strategies, P must, first of all, make some hypothesis about R's mind (her presumed beliefs and goals) by introducing some 'evidence' in one or more root nodes of the network. He then considers a candidate strategy to test and introduces in the network some evidence which corresponds to this strategy. The various pieces of evidence available are propagated in the network and their effect on the probability of the node (Int R Do(R, p)) is observed. If application of the selected strategy fails, the previous evidence is retracted and another candidate strategy is considered. Let us consider three examples which were produced by a simulation on the network in Figure 3. In the three cases, P presumes that R can eat vegetables: Prob(Bel R CanDo(R, p)) = 1. To follow these examples, please refer again to Figure 3.

Strategy S1: ***Acting on a non emotional goal, with cognitive activation of this goal***

P assumes that *losing weight* is of high value to R: Prob((V-Goal R LoseWeight)=High) = 1 and that R is convinced about the means-end relationship between *eating vegetables* and *losing weight:* Prob(Bel R ((Do(R,EatVegs)$\rightarrow$ $\Diamond$LoseWeight))=1 and between *losing weight* and *being in good health:* Prob(Bel R

[4] P may strengthen his persuasion strategy, thus increasing the likelihood for R to intend p, by either acting on several goals q_1 …. q_n (and showing how p is instrumental to each of them) or increasing the certainty of the means-end relationship between p and q_i (Poggi, 2004).

(LoseWeight→ ◊IsInGHealth))=1. Thus, P says: *Your cholesterol level is high! You should eat vegetables, to be in shape,* and presumes that R will believe him: Prob(Bel R HighChol) = 1. This activates the goal of *being in good health* (with prob = .8) and, consequently, of *losing weight* (prob = .66). However it does not produce a satisfying level of intention: Prob(Int R Do(R, p)) = .60.

Strategy S2: ***Acting on a non emotional goal, with emotional activation of this goal***

P assumes, again, that *losing weight* is of high value to R and that R is convinced about the means-end relationship between *eating vegetables* and *losing weight*. But now he says: *You are disgustingly fat! You should eat more vegetables:* Prob(Bel R IsDisgFat)=1. This arouses the emotion of shame in R, it activates her goal of *saving face* and generates the goal of *losing weight*, but with a higher probability than with strategy S1 (.85); consequently, the final level of intention produced is also higher (prob = .70).

Strategy S3: ***Acting on the goal to feel an emotion***

Noticing that R is very sad, P changes the goal of his persuasion strategy (from *to be in good health* to *to be in good mood*), and applies a strategy aimed at acting on the latter goal. He assumes, this time, that *being in good mood* is of high value to R: Prob((V-Goal R IsInGMood)=High) = 1. He says: *You look sad. Do you know that eating vegetables increases good mood?* This induces in R the belief about the means-end relation between *eating vegetables* and *being in good mood*: Prob(Bel R (Do(R, EatVegs)→ ◊IsInGMood))=1 and activates (with some uncertainty) the goal of *being in good mood*: Prob(A-Goal R IsInGMood)=.7 but produces a lower level of intention than with Strategy S2 (prob = .63).

In the conditions established in our simulation, strategy S2 seems to be the most convenient among those tested by P, with the presumed mental state of R. This is due to two conditions implicitly introduced in the network when we assigned the parameters (conditional independency tables) to it: i) the cognitive activation of the goal *to lose weight* is, to this recipient R, weaker than the emotional one; and ii) acting on the goal to feel the emotion of *being in good mood* is (again, to this recipient) a bit weaker than acting on the goal *to lose weight*.

5. Related Work

Our work builds on several research veins. A remarkable body of results about modelling of persuasion strategies in AI comes from Walton and Reed's research on argumentation schemes, the majority of which are focused on 'persuading to believe' (Walton, 1990; Walton and Reed, 2002). Among the few of them which are aimed at 'persuading to do', the most commonly applied and popular one is the *Argument from Consequences*. Our schemes of *generation of intentions,* either *by acting on 'non emotional' goals* or *by acting on the goal (not) to feel an emotion,* may be seen as particular cases of this argument. They are attempts at specifying some ways in which a particular course of action may have good (or bad) consequences, that is, may favour or thwart some goal. Other persuasion schemes suggested by the cited authors, such as the *Argument from waste* and the A*rgument from popular practice*, may be interpreted in terms of particular instantiations of our representation formalism:

a. The *argument from waste* leans on the belief that, when one has used some resources to achieve some goal, even if the goal has not yet been achieved, one should persist in this pursuit, to avoid wasting the resources employed so far (provided one thinks that the goal in question is still achievable). This is the basis of the *sunk cost effect* (e.g. Arkes and Blumer, 1985; Bazerman, 1998): for instance, the reason why we decide to keep waiting the bus after 30' wait, or to keep gambling in roulette after losing almost all of our money. Doing so

corresponds to pursuing the general meta-goal of *not wasting resources that have been already used.* To this regard, an example of persuasive strategy in the healthy eating domain might be: *You already spent a lot of money and effort to be in shape. Why don't you keep trying to lose weight?* Here, the *argument from waste* may be seen as a combination of goal activation (the goal being *to avoid wasting resources)* and intention generation (the intention being *to (continue to try to) lose weight*): thanks to this argument, R will 'remind' that in fact she has the goal of not wasting resources, with respect to which completing the 'task' of losing weight is a means.

b. The *argument from popular practice* belongs to a class of arguments about actions which are believed to be performed by all (or most of the) people. The argument appeals to the goal of *doing the same things other people do*. This argument assumes that: (i) if a large majority of people (everyone, nearly everyone etc) do something, they probably believe that doing this is right; and (ii) if something is generally considered as right, doing it corresponds to a prudent course of action. This kind of argumentation is, again, a particular case of the graph shown in Figure 1a, where the goal involved is 'PrudentCourseOfAction'.

As we anticipated in the Introduction, the interest for non purely cognitive aspects of persuasion is relatively recent. In his book on *The place of emotion in argument*, Walton (1992) claims that "Two factors combine to enhance the trickiness of arguments that appeal to emotion. One is that an appeal to emotion may not be *relevant*, meaning that it may not contribute to the goals of dialogue…. The other is that arguments based on emotional appeals tend to be *weak* arguments, based on presumptions rather than hard arguments….Such arguments become *fallacious* when the proponent exploits the impact of the appeal to disguise the weakness and/or irrelevance of the argument" (pp. 1-2). The author examines carefully some classical arguments (*argumentum ad populum, ad misericordiam, ad baculum* and *ad hominem*) to prove that emotional appeals are neither right or wrong in themselves, but should be known to both guard oneself against them and to use them appropriately. A few years ago, Marcu (2000) advocated that "one of the fallacies of the current study of perlocutions is in assuming the hearer to be a 'rational' agent". In analysing the factors affecting the structure of arguments, Sillince and Minors (1991) support with examples the idea that an argument is strong "if it matches the emotional expectations of the hearer, if it takes account of the emotional determinants behind the arguer's own position, if it 'gives a good feeling'". DAPHNE (Grasso, Cawsey and Jones, 2000) was one of the first operational experiences in this research area. In this system, 'values' and 'opinions' of the addressee are considered to select and justify arguments. While we attach a value to an agent's *goals*, values are attached, in DAPHNE, to a (*topic*, *perspective*) couple (where *topic* corresponds to our *plan* and *perspective* to our *goal*). For instance: *Being a vegetarian is good for health but may be bad for social life, when you go to dinner with non-vegetarians*. Guerini et al. (2003) start from the hypothesis that "persuasion is concerned with a-rational arguments" to propose a taxonomy of persuasion strategies and a rule-based meta-reasoning model which define how to select the strategy that best suits a given context; what they call the 'emotional' state of the addressee (to be *lively* or *tired* or *depressed*) is among the conditions of these meta-rules.

Our modelling formalism builds upon Rao and Georgeff's (1995) research on representation of mental attitudes in BDI agents. We add the 'emotion' component to the belief, desire and intention ones, thus going towards a BDI&E formalism. In addition, we extend the description of goal properties introduced by Cohen and Levesque (1990), who attached a *degree of persistence* to goals and defined intentions as a function of 'persistent' goals (those goals the agent will not give up until he thinks they will be satisfied or until he thinks they will never be true). We basically share Cohen and Levesque's view of intentions in terms of goals the agent chooses and is committed to pursue. In addition, we introduce other properties of goals such as:

- their being *active* or *inactive*, that is, included or not in the agent's 'goal balance', implying the agent's activity of assessment of their importance and feasibility (and possible comparison with other candidate goals) in view of their possible *translation* into intentions;

- their being *pre-existent* vs. *generated*, that is newly represented in the agent's mind;
- their having different degrees of *value* or importance to the agent.

Consideration of these properties is necessary in modelling persuasion strategies, either emotional and non emotional, as these strategies attempt to either activate R's inactive goals, generate in R some goals she did not have before or increase/decrease the value of her goals, so as to make R have some intention instrumental to those goals. However, in our view the use of such properties goes far beyond persuasion models: we assume they are in fact general properties of goals, which may enrich and make more dynamic the models of agents' mental attitudes.

All the formal systems stemming from Rao and Georgeff's (1995) and Cohen and Levesque's (1990) research in the domain of multi-agent systems aim at building a framework of agents' attitudes that, working within an appropriate architecture, is capable to represent their behaviours in a dynamic way. Our aim is rather to model the behaviour of a Persuader who reasons on the different possible ways to induce an intentional state in a Recipient: so our model is closer to user modelling research carried on within HCI or conversational systems. Here, among the theories proposed to represent the Agents' mental states, belief networks have gained more and more relevance. Actually, the advantage of representing persuasion as an intention formation process through a Bayesian network rather than through axiomatization of a logical theory is that it allows one to reason about the possible effects, on the Interlocutor's mind, of alternative persuasive strategies under uncertainty conditions. The Persuader's hypotheses about the Interlocutor's values, beliefs and propensity to feel emotions become evidence given as an input to the network. Evaluation of persuasive strategies becomes a 'what if' kind of reasoning whose goal is represented in terms of the Recipient's 'degree of conviction', that is, of the probability of his/her intention to do the wanted action. Considering probability theory and belief networks as a method for treating uncertainty is not a novelty in the argumentation community. BIAS was the first such system and is still being refined (Zuckerman, Jinah and McConachy, 2001). Gratton (2002) proposes to measure the *strength of support* in probabilistic terms and to estimate the effect of counterexamples against the argument in terms of this strength. Das (2002) measures probabilistically the *confidence* that the inference confers to an argument and proposes a method to semi-automatically aggregate individual arguments into belief networks, which aims at overcoming the well-known difficulty of building this complex kind of knowledge bases. Green (2003) applies a coding scheme based on a Bayesian network for describing arguments in medical genetics from a corpus of counselling letters, thus providing evidence that this formalism naturally applies to human argumentation messages. With our formalism, we unify different forms of persuasion to enable simulating how rational and a-rational forms may combine to produce an argumentation strategy which is suited to a particular context, that is to a Recipient with a given set of beliefs, goal values and personality traits. We show that belief networks are a knowledge representation and reasoning method that seizes several aspects of our theory.

6. Concluding Remarks

In this work we have presented a model of persuasion in terms of goals and beliefs that is viewed from the persuader's P perspective, thus focusing on his theory of the recipient's R mind, and P's planning strategies for influencing R, that is, for changing R's mental state so as to make her intend to do a certain action or plan. We have also tried to circumscribe the notion of persuasion in relation to such criteria as 'success' versus mere 'attempt' at persuasion, 'accidental' versus 'intentional', 'communicative' versus 'non communicative', 'manipulative' versus 'non manipulative', and 'coercive' versus 'non coercive' persuasion. Whereas some criteria – namely successful versus attempted, and manipulative versus non manipulative – in our view allow to discriminate between different kinds of persuasion, the remaining criteria allow to distinguish persuasion from

other, non persuasive, forms of influencing. That is, as already remarked, we place either accidental (unintentional) influencing, or non communicative influencing, or coercive influencing outside the 'realm' of persuasive strategies.

A qualifying feature of our model is the attempt to integrate emotional and non emotional persuasion. Emotional persuasion is just a sub-case of general persuasion. As its non emotional counterpart, emotional persuasion is aimed at generating, activating or increasing the strength of R's goals, so as to induce in R some intention instrumental to such goals. The specificity of emotional persuasion lies in the means used for accomplishing this task. That is, when using an 'emotional' strategy, P tries to generate, activate, etc. R's goals through the medium of R's emotions or R's beliefs and goals *about* her emotions.

We have identified two general modes of emotional persuasion: persuasion through actual arousal of emotions and persuasion through appeal to expected emotions. We have argued that the rational/irrational, as well as the argumentative/non argumentative dimension, do not allow to distinguish such forms of persuasion from the non emotional ones. Actually, one mode of emotional persuasion, the appeal to expected emotions, can be perfectly rational, as long as 'rational' implies the correct processing of the information available, the derivability of conclusions from premises, and the production of plausible means-ends relationships. An appeal to expected emotions is 'structurally' indistinguishable from any other 'argument from consequences' or, in our terms, 'intention generation by acting on pre-existing goals'. The only difference resides in the *content* of the goal on which P acts: in the appeal to expected emotions, this content is precisely that of 'feeling' a certain emotion rather than having a certain state of the world true. In fact, the content of a goal may regard either an external state of the world or an internal one, that is, a 'state of mind', be it a feeling (say, I may want to be cheerful), or a belief (I may want to believe in God), or even a goal (I may want to want others' welfare).

True, the other mode of emotional persuasion – persuasion through arousal of emotions – works very differently from the former, in that the aroused emotion can directly produce a certain goal, independent of R's reasoning and planning about means-ends relationship. Therefore, this is no doubt a form of non argumentative persuasion. However, in this context 'non argumentative' should be made equal to 'a-rational', rather than 'irrational' (as long as 'irrational' implies going *against* the dictates of reason). Moreover, it should be stressed that the direct production of a goal through emotional arousal is just one step which is generally included in a more complex persuasion strategy expecting a very 'rational' planning and behaviour on R's part. That is, once a certain goal is emotionally produced, R's reasoning and planning can be, and generally are, called into play in view of its achievement. Thus, even such a form of persuasion is partially based on the recipient's reasoning and planning abilities, which testifies to the constant mingling and intertwining of rational and a-rational ingredients in most persuasion strategies.

The language we propose for formalising rational and a-rational persuasion extends the classical components of an agent's mental state (beliefs, desires and intentions) with the attitude of 'emotion'. To render the dynamic process of mental state change which is typical of intention development, this language attributes a 'generation' state, a 'value' and a 'degree of activation' to the agent's goals. Uncertainty in knowledge of these second-order beliefs by the Persuader is conceptualised in terms of probability and is manipulated by Bayesian networks updating algorithms. This enables to explicitly represent the various hypotheses of independence among mental state components and to get a measure of the 'strength' of persuasion strategies which may be employed to compare alternatives, evaluate the impact of hypotheses about the Recipient and (in perspective) to repair to possible failures of a selected strategy.

Acknowledgements

This work was financed, in part, by HUMAINE, the European Human-Machine Interaction Network on Emotion (EC Contract 507422). We thank Valeria Carofiglio for contributing to build a prototype BN to model the persuasion strategies described in this paper.

References

Ajzen, I. 1985. From intentions to actions: A theory of planned behaviour. In J. Kuhl and J. Beckmann (Eds.), Action control: From cognition to behavior. Berlin: Springer-Verlag, pp. 11-39.

Arkes, H.R., and C. Blumer. 1985. The psychology of sunk cost. Organizational behavior and human decision processes 35:124-140.

Batra, R., and M.L.Ray. 1985. How advertising works at contact. In L. Alwitt and A.A. Mitchell (Eds.), Psychological processes and advertising effects: Theory, research, and application. Hillsdale, NJ: Erlbaum, pp. 13-43.

Bazerman, M. 1988. Judgement in managerial decision making. New York: Wiley.

Beisecker, T.D., and D.W. Parson. Introduction. In T. D. Beisecker and D. W. Parson (Eds.), The process of social influence. Englewood Cliffs, NJ: Prentice Hall, 1972, pp. 1-6.

Bell, J., and Z. Huang. 1997. Dynamic goal hierarchies. In J. Bell, Z. Huang, and S. Parsons (Eds.), Proceedings of the Second Workshop on Practical reasoning and rationality. Manchester, UK: University of Manchester, pp. 97-110.

Bratman, M.E. 1987 Intentions, plans, and practical reason. Cambridge, MA: Harvard University Press.

Burnell, P., and A. Reeve. 1984. Persuasion as a political concept. British Journal of Political Science, 14: 393-410.

Buss, D.M., R. Larsen, D. Westen, and J. Semmelroth. 1992. Sex differences in jealousy: Evolution, physiology, and psychology. Psychological Science, 3: 251-255.

Carofiglio, V. 2004. Modelling argumentation with belief networks. In G.Carenini, F.Grasso and C.Reed (Eds.) Workshop on Computational Models of Natural Arguments (CMNA'04), Valencia, August 2004.

Carofiglio, V., and F. de Rosis. 2003. Combining Logical with Emotional Reasoning in Natural Argumentation. In C. Lisetti, C. Conati and E. Hudlicka (Eds): Proc. of the UM'03 Workshop on Affect. Pittsburgh, 2003. pp.9-15

Carofiglio, V., F. de Rosis, and R. Grassano. 2005. Dynamic models of mixed emotion activation. In D. Canamero and R. Aylett (Eds): Animating expressive characters for social interactions. Amsterdam: John Benjamins.

Castelfranchi, C. 1990. Social Power: a missed point in DAI, MA and HCI. In Y.Demazeau and J.P.Mueller (Eds.), Decentralized AI. North-Holland: Elsevier, pp.49-62.

Castelfranchi, C. 1996. Reasons: belief support and goal dynamics. Mathware and Soft Computing 3: 233-247.

Castelfranchi, C. 2000. Affective appraisal versus cognitive evaluation in social emotions and interactions. In A.Paiva (Ed.): Affective Interactions. Springer: Berlin, pp.76-160.

Chaiken, S., W. Wood, and A.H. Eagly. 1996. Principles of persuasion. In E. T. Higgins and A. W. Kruglanski (Eds.), Social psychology: Handbook of basic principles. New York: Guilford, pp.702-742.

Cohen, P.R., and H.J. Levesque. 1990. Intention is choice with commitment. Artificial Intelligence 42: 213-261.

Conte, R., and C. Castelfranchi. 1995. Cognitive and social action. London: University College.

Daly, M., M. Wiston, and S.J. Weghorst. 1982. Male sexual jealousy. Ethology and Sociobiology, 3: 11-27.

Das, S. 2002. Logic of probabilistic arguments. In G.Carenini, F.Grasso and C.Reed (Eds): Workshop on Computational Models of Natural Arguments, in the scope of ECAI'02.

Fishbein, M, and I. Ajzen. 1975. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.

Frijda, N.H. 1986. The emotions. New York: Cambridge University Press.

Gordon, R.M. 1987. The structure of emotion. Cambridge: Cambridge University Press.

Grasso, F, A. Cawsey, and R. Jones. 2000. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. Int.J. Human-Computer Studies 53: 1077-1115.

Gratton, C. 2002. Counterexamples and degree of support. In G.Carenini, F.Grasso and C.Reed (Eds): Workshop on Computational Models of Natural Arguments, in the scope of ECAI'02.

Green, N. 2003. Towards an empirical model of argumentation in medical genetics. In G.Carenini, F.Grasso and C.Reed (Eds.), Workshop on Computational Models of Natural Arguments.

Guerini, M, O. Stock, and M. Zancanaro. 2003. Persuasion models for intelligent interfaces. In G.Carenini, F.Grasso and C.Reed (Eds.), Workshop on Computational Models of Natural Arguments.

Haddadi, A, and K. Sundermeyer. 1996. Belief-desire-intention agent architectures. In G.M. P. O'Hare and N.R. Jennings (Eds.), Foundations of distributed artificial intelligence. London: Wiley, pp. 169-185.

Jennings, N., S. Parsons, P. Noriega, and C.Sierra. 1998. On argumentation-based negotiation. In Proceedings of Int. Workshop on Multi-Agent Systems.

Jorgensen, P.F. 1998. Affect, persuasion, and communication processes. In P. A. Anderson and L. K. Guerrero (Eds.), Handbook of communication and emotion. San Diego, CA: Academic Press, pp. 403-422.

Kienpointner, M. 1992. How to Classify Arguments. In F.H. van Eermeren, R.Grootendorst, J.A.Blair, and C.A.Willard (Eds.), Argumentation Illuminated, pp.178-188.

Lazarus, R.S. 1991. Emotion and adaptation. New York: Oxford University Press.

Levy, D.A., B.E. Collins, and P.R.Nail. 1998. A new model of interpersonal influence characteristics. Journal of Social Behavior and Personality 13: 715-733.

Lisetti, C.L., and P. Gmytrasiewicz. 2002. Can a rational agent afford to be affectless? A formal approach. Applied Artificial Intelligence, Special Issue on 'Merging Cognition and Affect in HCI'. F de Rosis (Guest Editor), pp.7-8.

Marcu, D. 2000. Perlocutions: The Achille's heel of speech act theory. Journal of Pragmatics.

Miceli, M., and C. Castelfranchi. 2002. Emozioni. In C. Castelfranchi, F. Mancini, and M. Miceli (Eds.), Fondamenti di cognitivismo clinico. Torino, Italy: Bollati Boringhieri, pp. 96-129.

Miller, G.A., E. Galanter, and K.H.Pribram. 1960. Plans and the structure of behavior. New York: Holt.

O'Keefe, D.J. 2002. Persuasion: Theory and research. Thousand Oaks, CA: Sage.

Oatley, K., and P.N. Johnson-Laird. 1987. Towards a Cognitive Theory of Emotions. Cognition and Emotion: 29-50.

Ortony, A. 2003. On making believable emotional agents believable. In R. P. Trappl (Ed.), Emotions in humans and artefacts. Cambridge: MIT Press.

Pearl, J. 1988. Probabilistic Reasoning in Expert Systems: Networks of Plausible Reasoning. San Mateo, CA: Morgan Kaufmann (Pubs.).

Pears, D. 1984. Motivated irrationality. New York: Oxford University Press.

Perelman, C., and L. Olbrechts-Tyteca. 1960. The New Rhetoric: a Treatise on Argumentation. University of Notre Dame, IN: Notre Dame Press.

Petty, R.E, and J.T. Cacioppo. 1986. The Elaboration Likelihood Model of Persuasion. In L.Berkowitz (Ed.), Advances in Experimental Social Psychology. New York: Academic Press, 19, pp. 123-205.

Poggi, I. 2005. A goal and belief model of persuasion. Pragmatics and Cognition.

Poggi, I., and C. Pelachaud. 2000. Performative facial expressions in Animated Faces. In J.Cassell, J.Sullivan, S. Prevost, and E.Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: The MIT Press, pp.155-188.

Rao, A.S., and M.P. Georgeff. 1995. BDI Agents: from theory to practice. Proceedings of the First international Conference

of Multiagent Systems (ICMAS-95), San Francisco, USA, 1995.

Rosenblueth, A., N.Wiener, and J. Bigelow. 1968. Behavior, purpose and teleology. In W. Bucker (Ed.), Modern system research for the behavioral scientist. Chicago: Aldine.

Rosselli, F., J. Skelly, and D.M. Mackie. 1995. Processing rational and emotional messages: The cognitive and affective mediation of persuasion. Journal of Experimental Social Psychology 31: 163-190.

Rowe, G.W.A., C.A. Reed, and J. Katzav. 2003. Araucaria: Marking-up arguments. In Working notes of the European Conference on Computing and Philosophy, Glasgow, 2003.

Schwarz, N. 1990. Feelings as information: Informational and motivational functions of affective states. In E.T. Higgins and R.M. Sorrentino (Eds.), Handbook of motivation and cognition: Foundations of social behavior. New York: Guilford Press. Vol. 2, pp. 527-561.

Sillince, J.A.A., and R.H. Minors. 1991. What makes a strong argument? Emotions, highly-placed values and role playing. Communication and Cognition 24, 3-4: 281-298.

Spiegelhalter, D.J. 1986. Probabilistic Reasoning in L.N.Kanal and J.F.Lemmer (Eds.), Predictive Expert System, Uncertainty in Artificial Intelligence. North-Holland: Elsevier Science, 4, pp. 357-369.

Symons, D. 1979. The evolution of human sexuality. New York: Oxford University Press.

Tooby, J., and L. Cosmides. 1990. The past explains the present: Emotional adaptations and the structure of ancestral environment. Ethology and Sociobiology 11: 375-424.

Toulmin, S. 1958. The Use of Arguments. Cambridge, Ma: Cambridge University Press.

Walton, D.N. 1990. What is Reasoning? What is Argument? The journal of Philosophy 87: 399-419.

Walton, D.N. 1991. Nonfallacious Arguments From Ignorance. American Philosophical Quarterly 29(4): 381-387.

Walton, D.N. 1992. The place of emotion in argument. Philadelphia: The Pennsylvania State University Press.

Walton, D.N. 1996. Practical Reasoning and the Structure of Fear Appeal Arguments, In Philosophy and Rhetoric, 1996. The Pennsylvania State University, University Park. Vol. 29.

Walton, D.N. 2000. The Place of Dialogue Theory in Logic, Computer Science and Communication Studies. In Syntheses 2000 Netherlands: Kluwer Academic (Publ.). 123, pp. 327-346.

Walton, D.N., and C.A.Reed. 2002. Argumentation schemes and defeasible inferences. In G.Carenini, F. Grasso, and C.A. Reed (Eds.),Working Notes of the ECAI'2002 Workshop on Computational Models of Natural Argument, Lyon, 2002, pp.45-55.

Zukerman, I., N. Jinah, R. McConachy, and S.George. 2001. Recognizing intentions from rejoinders in a bayesian interactive argumentation system. In PRICAI2000, Melbourne, Australia.

Zukerman, I., and D.W.Albrecht. 2001. Predictive statistical models for user modeling. User Modeling and User-Adapted Interaction 11: 1-2.

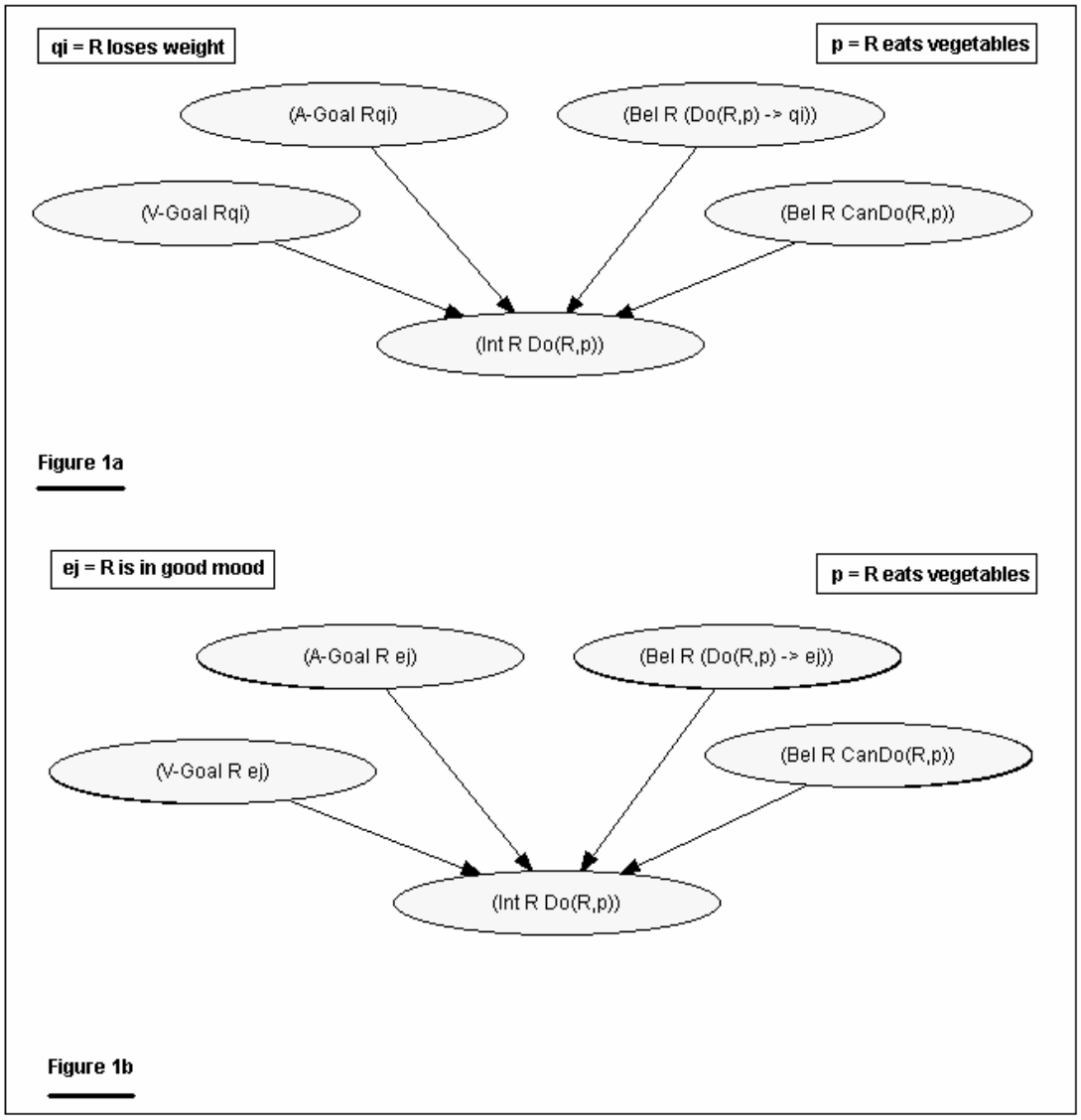

Figure 1: Generation of intentions by acting on 'non emotional' (2a) or 'emotional' (2b) goals

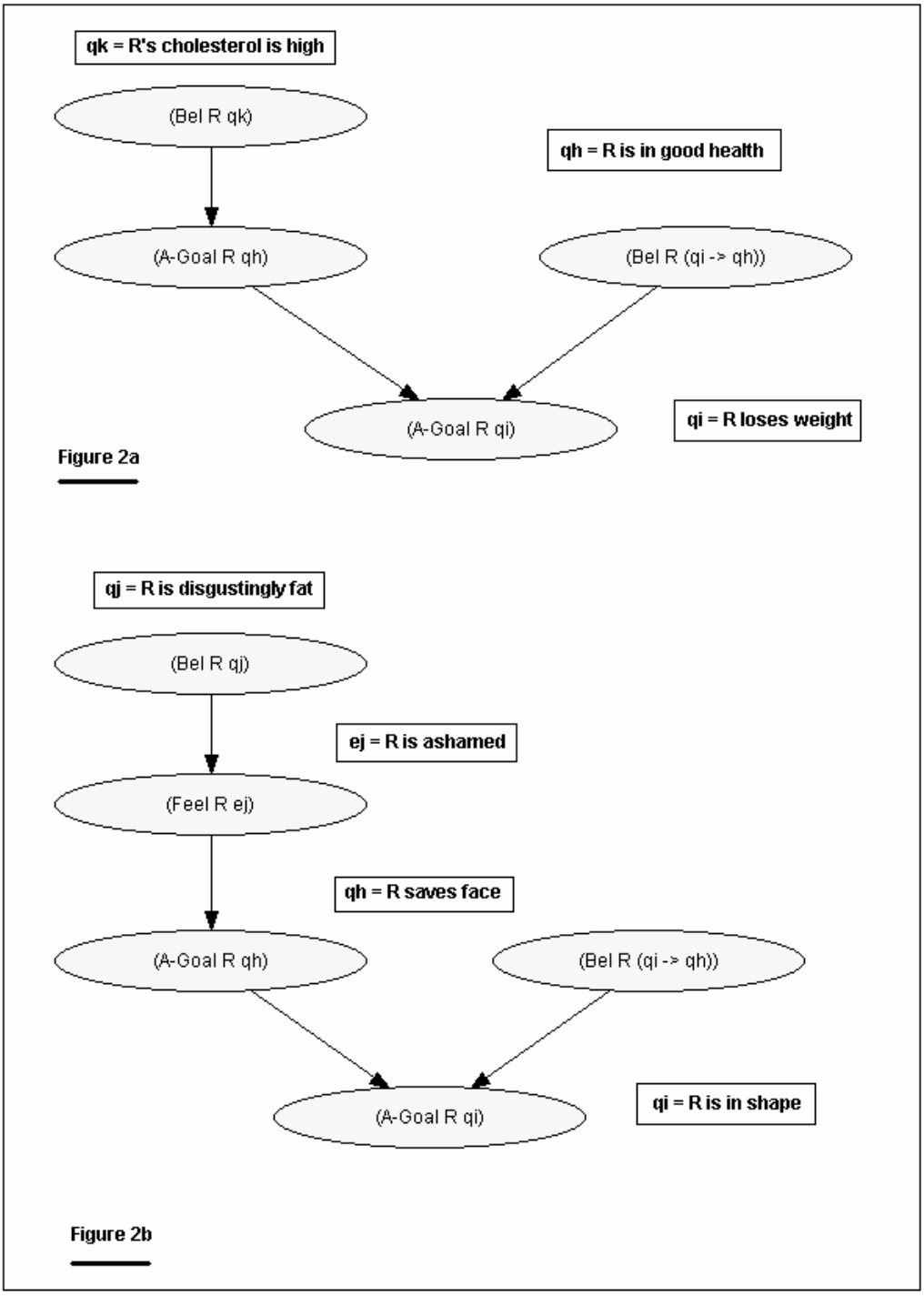

Figure 2: Cognitive (3a) and emotional (3b) activation of goals

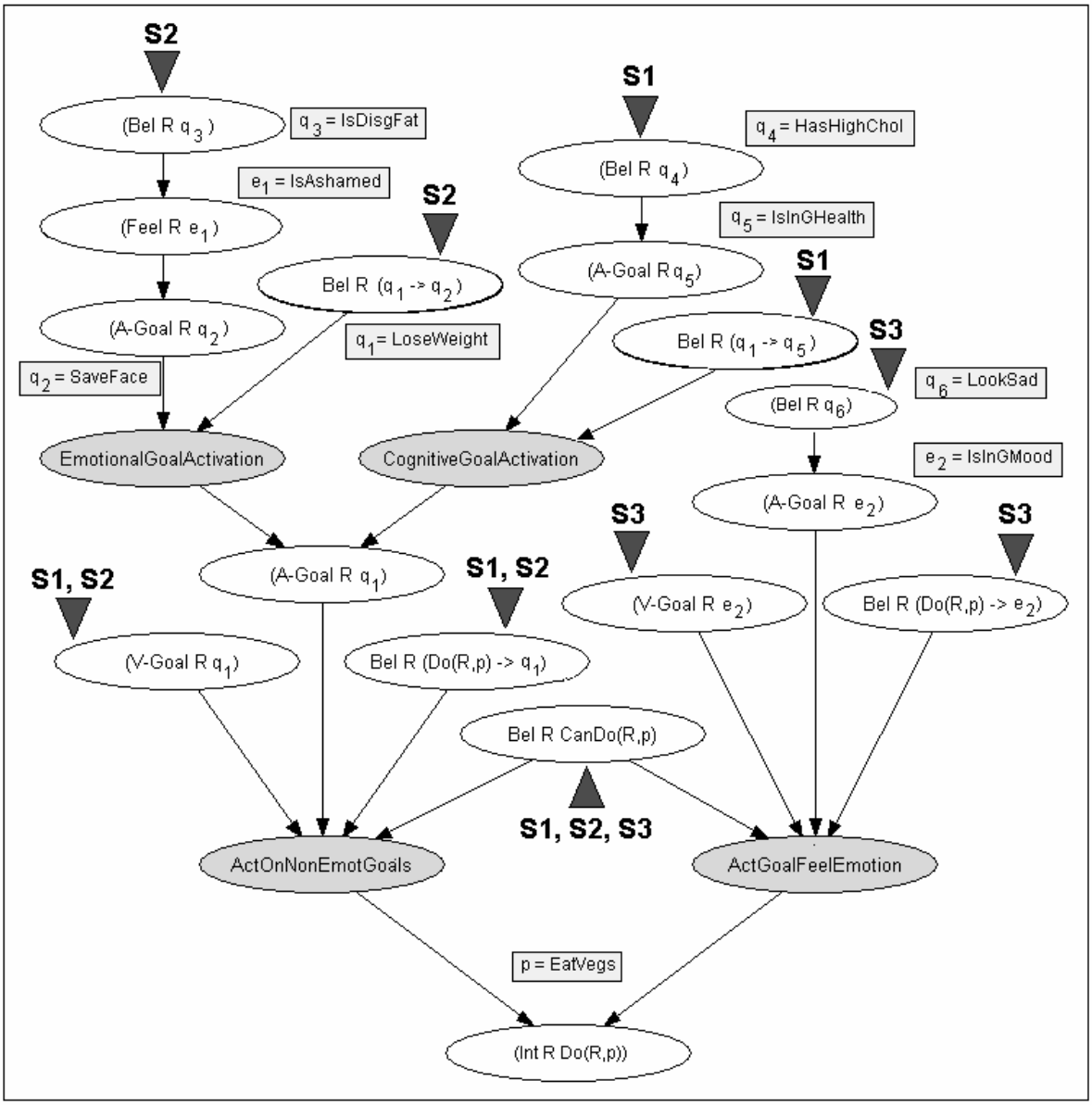

Figure 3: Combined persuasion strategies