Influence of FFM/NEO PI-R personality traits on the rational process of autonomous agents

François Bouchet [a], Jean Paul Sansonnet [b]
[a] *SMART Laboratory, McGill University, 3700 McTavish, Montreal, Canada*
E-mail: francois.bouchet@mcgill.ca
[b] *LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France*
E-mail: jps@limsi.fr

Abstract. In this paper, we present an approach based on the principle that psychological capacities, especially personality traits, influence the decision making process of rational agents. Using a three-level (trait, facet, scheme) extension of the FFM/NEO PI-R taxonomy facilitating its computational implementation, we propose a model for the expression of personality traits in terms of so-called influence operators that add meta control rules to the cycle of rational BDI agents. We distinguish eight different classes of influence operators, depending on the step of the deliberation cycle that they influence and on the operand to which they are applied (goals, actions, plans or functions). Concrete examples are provided through a complete definition of the operators necessary to express one of the personality traits of the Five Factor Model, Conscientiousness, resulting in a set of 30 operators. Finally, we discuss the way to design a specific character through the use of an activation matrix, providing values for each scheme and operator.

Keywords: Agent cognitive modeling, Computational personality, Personality traits, Rational and Psychological agents.

1. Introduction

1.1. Context

This work is situated in the framework of recent studies on cognitive modeling for conversational agents. Agent cognitive modeling stems from works in Artificial Intelligence (AI) focusing on reasoning about actions and plans of artificial agents. The main objective of those rational agents is to achieve behaviors that are *effective* in order to reach their goal. However, in the particular context of conversational agents, those actions and plans mainly aim at handling multimodal interactions with human users (usually, people of the general public), and it is then crucial for the agents to provide a realistic emulation of human cognitive functions to be perceived as *humanly believable*. Therefore, psychological notions, scarcely addressed in traditional artificial intelligence community studying rational agents, are now brought as a top priority in scientific concerns.

Although presented above as clearly separated notions, rational reasoning and psychological reasoning capacities actually work in a somewhat intricate way [20,24]. For example, some factors linked with high level psychological functions, such as personality traits [32] (usually stable and durable), could be integrated into existing architectures of rational agents to turn them into conversational agents that would be both rationally efficient and humanly believable.

With this objective in mind, from previous work on assistant agents [52,57,6], we have come to propose a conceptual framework dedicated to the computational study of the nature of the relationships between the rational and the psychological capacities of an agent [53]: namely, the R&B architecture (standing for *rational and behavioral*[1], where *behavioral* refers exclusively to psychological behaviors). R&B is a generic framework enabling the computational defi-

[1] `http://www.limsi.fr/~jps/research/rnb/rnb.htm`

nition and experimentation of various hypotheses that researchers may bring forth about the relationships between rational reasoning and psychological reasoning.

In the course of recent works in the R&B framework, we have proposed an approach in which psychological phenomena, such as personality traits, are implemented in terms of *influence operators*, intended to control (or even sometimes to alter) the rational reasoning process of an artificial agent. For the time being, this approach has been put to the test on a arbitrarily selected restricted set of psychological phenomena (*e.g.* behaviors related to traits such as lazy, gloomy or cheerful) [7,56], thus providing a good illustration of the underlying principles of the R&B framework and proving the feasibility of the implementation of some psychological phenomena in it. Nonetheless, to become a truly useful testbed for agent psychological reasoning, this framework requires to prove that it can cover a significant proportion of the psychological behaviors of a person, such as personality traits. An illustration of this key issue is proposed in Table 3 and discussed in details in section 2.5.
For these reasons, there is now a real need for defining a domain of psychological phenomena to take into consideration, and for treating this whole domain (as opposed to a set of separated case studies) in a generic and computationally tractable way, in order to lead to effective implementations.

1.2. Methodology

In this context, our methodology for applying the R&B principles upon a significant domain of personality traits is illustrated in Figure 1, and rely on three main steps: building a taxonomy of personality traits based on the literature that would be rich enough to be computationally tractable, proposing a formal model supporting operators of influence during the rational reasoning, and finally, linking those two elements through the use of an activation matrix.

1.2.1. Building a computationally tractable enriched taxonomy of personality traits

In order to propose computational models for psychological phenomena, computer scientists can start from works reported in the literature on psychology, which provide several taxonomies describing human personality [42]. Stemming from state of the art works in psychology, we have considered the popular Five Factor Model (FFM) taxonomy [27], which is based upon five main psychological dimensions, called *traits*. FFM has been refined by Costa and McCrae [13] in a two-level taxonomy of psychological concepts called *facets* (with six facets associated to each trait), called FFM/NEO PI-R in the literature.

However, these taxonomies are too generic for a direct computational implementation, and it is hence necessary to complete and to adapt them in order to facilitate their integration into effective computer systems. This is the reason why we have extended the FFM/NEO PI-R taxonomy with a third sub-level of concepts called *behavioral schemes* (abbreviated as *schemes*) hence issuing in an enriched taxonomy called FFM/NEO PI-R/BS, which is intended to provide instantiations of actual psychological behaviors associated with the FFM/NEO PI-R categories. A scheme is composed of a semantically consistent set of glosses describing the actual behavior of a person that exhibits the scheme in their personality profile. The detailed synthesis process of the FFM/NEO PI-R/BS taxonomy is described in [54] and the results are available as a free resource online, hence we will only briefly summarize it in Section 2.3.

1.2.2. Building a formal model supporting influence operators during the rational reasoning

State of the art works in AI and Multi-agent Systems (MAS) make large use of the BDI framework [48], where agents perform rational reasoning over concepts such as Beliefs, Desires and Intentions, and are controlled through the so-called BDI *deliberation cycle*. Given a typical BDI agent model, our first issue is to substantiate the R&B claim that it can be influenced by meta-heuristics implementing personality traits. This entails that the deliberation cycle should exhibit capabilities for being influenced, according to various modalities. It led to the elicitation of eight distinct classes of influence operators, which add meta control over the deliberation cycle upon actions and plans.

1.2.3. Linking personality to influence operators through an activation matrix

Given the FFM/NEO PI-R/BS taxonomy and the eight classes of influence operators, we have to: a) define a psychological agent model; and b) show how this psychological model can be implemented through influence operators.

First one must emphasize that by 'psychological agent' we don't mean a software program capable of performing the same complex psychological processes pertaining to human beings (say a program experiencing internal emotional states). Rather we will use the term 'psychological agent' to denote a software

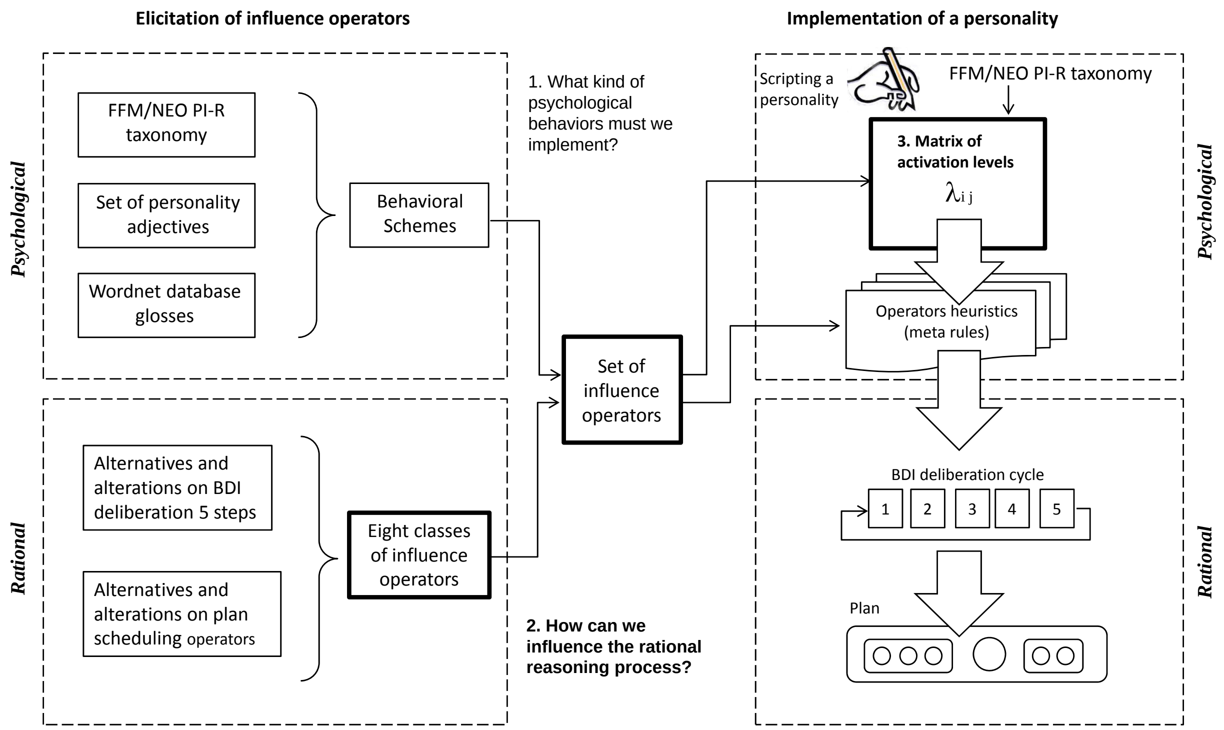

Fig. 1. Global methodological framework. Parts discussed in the paper are in **bold**: 2) elicitation of influence operators and 3) principles of activation matrix.

program whose observable behavior (actions in/on the world, expressions and gestures, *etc.*) make it possible for human subjects to qualify it with adjectives/adverbs usually employed in the description of the personality traits of a person: *lazy*, *shy*, *friendly*, *truthful*, and hundred of others.

For this study, we attempt to sketch a generic psychological agent model that is a model that covers a large class of psychological phenomena. As this can quickly prove to be too large of an issue, we state four important restrictions:
1. Restriction to the static part of psychological phenomena *i.e.* the computational expression of the personality traits (it will be based on the well-used FFM/NEO PI-R/BS taxonomy). Emotions will therefore be considered out of scope.
2. Restriction to static handling of traits, meaning that we do not address the evolution of personality traits, for example over life. Outside of the particular context of companion agents, meant to interact with users over a very long period of time, it is a reasonable assumption as stability over time is the key difference between personality traits and emotions.
3. Restriction to a generic definition of the personality traits, meaning that no context is currently treated in our model. A drawback of this restriction is that it can make difficult the handling of pragmatic-sensitive notions such as irony.
4. Restriction to the description of a formal model, meaning that we do not address, in this particular paper, issues relative to evaluation and validation, such as examples of scenarios for validation and possible metrics for evaluation.

Second we propose to establish a declarative implementation of FFM/NEO PI-R/BS schemes in terms of influence operators based on a single data structure: a matrix of activation levels.

Outline of the paper: The rest of this paper is organized as follows: in Section 2, we start by presenting the general context of the research: motivations for a generic approach to rational and psychological agents, related works implementing psychological phenomena into BDI architectures, and we provide a summary of our previous work about the enrichment of FFM/NEO PI-R taxonomy into FFM/NEO PI-R/BS, which grounds this study. In Section 3, a formal model is introduced to support the rational dimension of the agent, which is then used in Section 4 to elicit eight classes of influence operators that can be applied to sub determined plans. Finally, in Section 5, the principles of the activation matrix are presented and illustrated by the implementation of the schemes pertaining to FFM/NEO PI-R/BS trait Conscientiousness. We conclude in Section 6 with discussing possible future works to go beyond this paper aforementioned restrictions.

2. Context of the research

2.1. Definition of agent categories

1. Rational Agents: For many years now, the notion of rational agent has been developed over several successful AI frameworks. Several rational agent architectures are based on practical reasoning, in the following of the SOAR [33] and ACT-R [3] frameworks, or more recently BDI agents [10,48]. In Bratman's theory of practical reasoning, an agent's behavior is modeled by specifying *beliefs*, *goals* and *plans* and is effectively produced through the agent's so-called deliberation cycle. These architectures have been quite successful at creating both autonomous and multi-agent systems that are *competent*, *i.e.* capable of planning and performing automatic tasks without human interaction or sometimes under the control of experts (*e.g.* human supervisors of a system).

2. Conversational Agents: With the development of new Internet technologies, agents have been more and more in interaction with a large class of people of the general public, rather than experts, as revealed by the growth of so-called conversational agents [11]. In this new context, authors have claimed agents should be both *competent* (thanks to their symbolic reasoning capacities [51]) and psychologically *relevant*, in order to increase: a) the acceptability factor [16], especially when they deal with people of the general public, b) the efficiency factor or usability, in the sense of Nielsen [44] and Dillon [17], within the agent/human interactional process itself (*e.g.* a transaction process, a teaching process…).

3. Assistant Agents: Requirements raised by conversational agents entail the definition of a new class of agents, which are able to provide ordinary users with an intelligent mediation with a computer system: so-called assistant agents [35]. We will use the word 'assistant' as a generic term to embrace several sub cases of *conversational situations*[2] where assistant agents can endorse various roles: *presenter agents*: non interactive entities just displaying informational content about a topic; *help agents*: software tools providing assistance to users in need for help to fulfill a task with an unfamiliar computer application; *butler agents*: personal assistants as described by Maes [35]; *friend agents*: mates/friends in chat and social activities (*e.g.* chatbots, social networks, *etc.*); *companion agents*: partners/opponents in games, role-based simulations or serious games; *teacher agents*: tutors/teachers in learning activity; *training agents*: instructors (*e.g.* in training simulations); *coach agents*: monitor and control personal life (*e.g.* for a diet).

4. Psychological Agents: since the introduction of the notion of *believable agents* in the mid-1990s [4,36,29], there have been various attempts at implementing psychological features into cognitive architectures. For example, the works of Rousseau and Hayes-Roth [50] have established a first ground by providing examples of how personality factors (also called personality traits) can be implemented into the cognitive architecture of a rational artificial agent. All these authors emphasize the fact that agents should exhibit two reasoning engines: a rational one, but also a psychological one.

2.2. Definition of the psychological domain

2.2.1. Psychological frameworks

Researchers have proposed many theoretical frameworks (each with sub categorizations) to the psychology of a person [38], of which four prominent are listed in historical order:

1. Psychoanalysis by Sigmund Freud, which is a psychodynamic perspective (*i.e.* a notion relative to personality development) emphasizing unconscious motivation. Although intended for therapy, it is also used as a theory of personality.

2. Types and traits by many authors discussed further, which is a dispositional perspective (*i.e.* a notion rel-

[2] In this study, we restrict the field of the function of assistance to the situations where the user is a novice (*i.e.* a person of the general public) and where assistance tools are explicitly called by the user.

Table 1
The two-level FFM/NEO PI-R taxonomy

FFM Traits	6 FFM/NEO PI-R facets for each trait
Openness	Fantasy, Aesthetics, Feelings, Actions, Ideas, Values
Conscient-iousness	Competence, Orderliness, Dutifulness, Achievement-striving, Self-discipline, Deliberation
Extraversion	Warmth, Gregariousness, Assertiveness, Activity, Excitement-seeking, Positive-emotions
Agreability	Trust, Straightforwardness, Altruism, Compliance, Modesty, Tender-mindedness
Neuroticism	Anxiety, Angry-Hostility, Depression, Self-consciousness, Impulsiveness, Vulnerability

ative to observed static behaviors), where types label each person as a single class while traits identify the degree to which several different personality factors are occurring within an individual; traits therefore provide a system of independent dimensions.

3. *Humanistic* by Abraham Maslow and Karl Rogers, is opposed to both psychoanalytic and type/trait theories and based on the study of healthy individuals and introduced two main notions : defense maneuvers (also called coping) and scale of needs (physiological, safety, belonging, esteem, self-actualization).

4. *Social-cognitive* by Albert Bandura. It is an optimistic view that allows growth and change based on person-centered therapy: the behavior of a person emerges from the interplay between this person and their socio-cognitive environment (*i.e.* a notion of reciprocal determinism).

Each framework can be considered in its own right for integration into a computer system. For example, in previous works influenced by the psychoanalytic school, we have proposed a model to implement so-called *cognitive biases* into a rational agent [55], mimicking by some aspects the concept of unconscious motivation from the Freudian theories. However, most of modern computer studies are based on the domain of personality traits and we will focus on this domain as well.

2.2.2. *Personality traits taxonomies*

When one is interested in the computational implementation of psychological phenomena, especially those related to personality traits, the most successful paradigm to day is the Five Factor Model (FFM), stemming from a convergent research from many authors in psychology during the last two decades. FFM is based upon five main psychological dimensions, often called OCEAN traits (standing for personality traits: Openness, Conscientiousness, Extroversion, Agreeableness and Neuroticism). FFM has taken upon Cattell's classification into 16 factors (16PF) [12]; the 16PF approach was still prominent in the 1980s and was supported by Eysenck's Personality Questionnaires (EPQ) [22], which are questionnaires (generally composed of yes/no questions) designed to assess the personality traits of a person.

Another approach to the taxonomy of traits is based on natural language, and more precisely, lexical resources such as the glosses found in dictionaries. Historically, the lexical hypothesis states that most of socially relevant and salient personality characteristics have become encoded in the natural language [2]. More recently, this approach has been promoted by many authors, such as Goldberg [26], and contributed to the definition of the FFM. For this reason, the work presented here is based on lexical resources and more precisely on lexical semantics.

2.2.3. *Refining traits with facets*

However, classifications such as Cattell's 16 factors or FFM's 5 factors remain very generic, which led some authors to propose to refine these taxonomies with facets that are bipolar (the concept and its antonym) positions. One of the most commonly used is the FFM/NEO PI-R taxonomy, introduced by Costa and McCrae [39] and summarized in Table 1 defines 6 facets for each factor in FFM, thus resulting in a more precise two-level taxonomy (traits/facets) of 30 bipolar positions.

2.3. *Refining* FFM/NEO PI-R *taxonomy with schemes*

Nonetheless, the taxonomy FFM/NEO PI-R is still too generic from a strict computational perspective because implementing psychological behaviors in software agents requires operational definitions: *e.g.* position CONSCIENTIOUSNESS/Orderliness is associated with the very general gloss: 'personal organization'. Adding operational glosses such as (taken from Table 2): 'concise and succinct', 'showing care in execution', 'Do things according to a plan', 'Likes to tidy up' *etc.* provides actual instantiations that can be more easy to implement from a computational perspective.

Actually, for a given facet, associated glosses can be categorized into sub sets describing a consistent behavior : for example, aforementioned glosses 'showing

Table 2

Excerpt from the Web resource on schemes in FFM/NEO PI-R/BS.

TRAIT: CONSCIENTIOUSNESS			
Facet 1: Competence			
Scheme: ASSUREDNESS			
	pole+ ASSURED		pole- INSECURE
N	**Gloss/Qitem**	**N**	**Gloss/Qitem**
366[a]	showing poise and confidence in your own worth	136	lacking self-confidence
479	having or marked by confidence or assurance	230	showing fear and lack of confidence
947	marked by excessive confidence	680	lacking or indicating lack of confidence or assurance
1290	marked by excessive complacency or self-satisfaction		
Q64[b]	Am sure of my ground		
Scheme: DECISIVENESS			
	pole+ DECISIVE		pole- HESITANT
478	persuaded of; very sure	1106	having or feeling no doubt or uncertainty; confident and assured
477	not liable to error in judgment or action	790	unsettled in mind or opinion
265	characterized by decision and firmness	791	marked by or given to doubt
Q61	Complete tasks successfully	856	lacking decisiveness of character; unable to act or decide quickly or firmly
Q62	Excel in what I do	1203	uncertain how to act or proceed
Q63	Handle tasks smoothly	Q67	Misjudge situations
Q65	Come up with good solutions	Q68	Don't understand things
Q66	Know how to get things done	Q69	Have little to contribute
		Q70	Don't see the consequences of things
		Q100	Postpone decisions
Facet 2: Orderliness			
Scheme: METHODICALNESS			
	pole+ METHODICAL		pole- DISORGANIZED
1100	concise and succinct	1107	completely unordered and unpredictable [..]
1249	methodical and efficient in arrangement or function	1143	lacking order or methodical arrangement or function
278	being effective without wasting time or effort or expense	1224	wildly disordered
598	showing care in execution	Q80	Am not bothered by disorder
Q75	Do things according to a plan		
Scheme: TIDYNESS			
	pole+ TIDY		pole- MESSY
597	clean or organized	375	lacking neatness or order
1035	marked by order and cleanliness in appearance or habits	918	dirty and disorderly
Q71	Like order	Q76	Often forget to put things back in their proper place
Q72	Like to tidy up	Q77	Leave a mess in my room
Q73	Want everything to be 'just right'	Q78	Leave my belongings around
Q74	Love order and regularity	Q79	Am not bothered by messy people

Other facets are: 3: Dutifulness, 4: Self-discipline, 5: Achievement-striving, 6: Deliberation.

[a] references to WordNet glosses [b] Q_i are references to Goldberg's Q-items.

care in execution', and 'Do things according to a plan', can be placed in the same set, say METHODICAL, while 'Likes to tidy up' should be placed in another set, say TIDY. This is the reason why, for each FFM/NEO PI-R position, glosses and items have been clusterized in so-called *behavioral schemes* (in short schemes) associated with congruent operational behaviors (an excerpt is given in Table 2 for trait Conscientiousness). Schemes and their related taxonomy FFM/NEO PI-R/BS are defined and fully presented in [54].

2.3.1. Scheme structure

Formally, each scheme is a data structure of the form:

$$\sigma \doteq \langle S_i, T_\varphi, \gamma_i, S_i^+[\tau_1^+, ..\tau_n^+], S_i^-[\tau_1^-, ..\tau_m^-]\rangle$$

where:
S_i is a symbol, unique identifier of the scheme σ;
T_φ is the category of the scheme in FFM/NEO PI-R, with T being a trait and φ a facet of that trait;
γ_i is the unique textual definition of the scheme;
S_i^+ (*resp.* S_i^-) is a symbol, unique identifier of the positive (*resp.* negative) pole;
$\tau_i^+, i \in [1,n]$ (*resp.* $\tau_j^+, j \in [1,m]$) are glosses associated to the scheme for the positive (*resp.* negative) pole.

Traditionally, facets are organized into two poles: the concept (+) and its antonym (-). Like facets, schemes are organized into poles but we use a three valued scale: tendency to exhibit the concept more than average (+), average (=), tendency to exhibit the concept less than average (-). Actually no glosses are attached to the average position because it denotes the behavior of an "average" person (defined in Section 5.2), while poles + and - denote behaviors that exhibits a divergence from average, which are the one encoded in the language to describe one's personality. Consequently, in Table 2, only poles + and - are displayed.

2.3.2. Scheme semantics

Technically, the semantics of a scheme σ is defined by two notions:
– the list of all the glosses τ associated to the positive and negative poles of σ;
– the position of σ into FFM/NEO PI-R, which makes it inherit of the semantics associated by psychologists to that particular position.

2.4. Towards a generic approach to the integration of personality into agents architectures

2.4.1. BDI agents

As mentioned previously, today outstanding approach in AI and MAS is the BDI framework from Bratman (1990) and Rao and Georgeff (1995), which focuses on reasoning about actions and plans [1] of agents situated in a given environment. The BDI approach has produced many operational platforms that are based on various formal tools such as logics (KGP, MINERVA, AGENTSPEAK, CAN) using rules (2APL), temporal logics (MetateM, the Golog family) or more straightforward programming, for example using Java (JACK, JASON, JADEX, JADE).

All these architectures are based on the so-called BDI deliberation cycle of the agent, which is traditionally decomposed in five main steps: Deliberation, Planning, Decision, Execution, Evaluation. A formal development of these five steps is presented in [48] or [58]. The BDI deliberation cycle plays a central role in the R&B architecture because influence operators apply to the five steps of this cycle, along eight modes that are discussed in Section 3.1.

2.4.2. Psychological studies on BDI agents

For some years now, the definition of agent cognitive models that are more complete because they integrate psychological notions and the definition of artificial agents that can interact in a more "human-like" manner with ordinary people has prompted the integration of some psychological phenomena into already existing AI architectures. For example, Gratch and Marsella [28] have implemented a model of emotions based onto a traditional SOAR architecture. Their works showed that SOAR architecture does not easily support the implementation of psychological behaviors and that the SOAR core engine must be adapted.

On the contrary, BDI architectures seem to offer an open and flexible engine (the deliberation cycle). For example, one can rely on frameworks like 2APL, which is a BDI based agent oriented programming language developed by Dastani [14] providing practical programming constructs to allow the generation, repair, and (different modes of) execution of plans based on beliefs, goals, and events. Emotions have also been incorporated as in [15]. Some authors have proposed improvements of BDI architectures exhibiting both rational reasoning modules and psychological reasoning modules [34]. For example, the eBDI model of Jiang *et al.* [31] implements emotions in a BDI architecture.

Other work on BDI and emotions, related to logic consideration includes [60].

One must also mention some closely related studies where authors focus on the influence of physical or cognitive capacities over the deliberation cycle, but not on actual psychological capacities, like moods or traits. For example, using the well-used agent creation platform JACK [30] that implements the BDI theory, CoJACK [21] is an extension layer intended to simulate physiological human constraints like the duration taken for cognition, working memory limitations (*e.g.* loosing a belief" if the activation is low or "forgetting the next step" of a procedure), fuzzy retrieval of beliefs, limited focus of attention or the use of moderators to alter cognition. A similar approach is taken for conversational agents in PMFserv [59].

2.4.3. *Conversational agents*

Closer to our work, Malatesta *et al.* [37] use traits to create different expressions of behaviors, especially by influencing the appraisal part of the OCC theory [46]. The difference is that their work focuses on how agents evaluate the results of their actions and of external events, not on the way they perform a task. Nonetheless, the idea that traits can differentiate agents' behaviors underlies this work. Also works of van der Weide *et al.* addresses the issue of personality traits in relation with practical reasoning [61], and [41] focus on reliability and helpfulness (but from a pure rational point of view, as in studies on non normative behaviors in agents [19]). One can also mention [25] and [43] which consider influences from a social point of view, and [5] for influences about output modalities in ambient environments.

In the same way, Rizzo *et al.* [49] have shown that goals and plans can be used to represent a character's personality in an efficient way, by attributing specific behaviors to the pursuit of each goal. Personality traits are used to choose between the multiple goals of a BDI agent (*i.e.* traits influence Desires). Once chosen, goals are planned and executed directly, whereas in our case, traits operate on already planned goals (*i.e.* traits influence Intentions). This applies also to [40], based on the well-used architecture of conversational agent GRETA [47], which involves models of personality for the expression of emotions (face, gesture, *etc.*), as well as to the FATIMA architecture [18] (stemming from [47]), which implements personality traits.

Table 3

Heuristic implementing a simple behavioral reaction.

Heuristic: Handling negative remarks to the agent

1: H_b: dislike-agent[DISLIKE$_{user}$[agent]] $\doteq$
2: $\to$ { $Q[i,$ `MAP` [energy, $\lambda\ x.x*0.9$]],
3: $Q[j,$ `MAP` [confidence, $\lambda\ x.x*0.9$]],
4: $Q[k,$ `MAP` [cooperation, $\lambda\ x.x*0.9$]]
5: }
6: $Q_i^+ \wedge Q_i$.value < -0.5 $\to$
`ADD` [$\mathscr{R}$.reply, TELL$_a$[energy, "tired"]]
7: $Q_j^+ \wedge Q_j$.value < -0.5 $\to$
`ADD` [$\mathscr{R}$.reply, TELL$_a$[confidence, "depressed"]]
8: }

Explanations:

This heuristic is composed of three guarded rules in `2:`, `6:` and `7:`

2: the agent performs a sequence of three internal queries, which update its mental model. 'energy' represents an intensity level of a variable in the mental model of the agent. $\lambda x.x*0.9$ is a λ-expression decreasing x by 10%.

6: when agent's energy is very low, the phrase "I feel tired" is appended to the list of things-to-say.

7: idem for variable 'confidence'; the rule scheduler ensures simultaneous expression of the two feelings: "I feel tired *and* depressed".

2.5. *The* R&B *framework*

In previous work, we have proposed the R&B framework [7,55] as an architecture dedicated to the study of the relationships between rational and psychological capacities, and which shares with aforementioned architectures the two motivations of psychological modeling and human-likeness. Actually, R&B also shares with those architectures an important drawback: its *adhocity*, in the sense that they are all based on separate case studies, each of them corresponding to a specific psychological phenomenon which is treated individually.

In order to illustrate how this adhocity can be a key issue, we give the example of a R&B heuristic adapted from [55] that handles a specific psychological phenomenon while making use of rules influencing the rational decision making process of an agent. Let us consider a user that puts a textual remark to a conversational agent: "I hate you!", transformed by the natural language processing module into a formal request in the form of: DISLIKE$_u$[agent]. We have implemented a possible reaction for this class of requests, which is described in Table 3. Although details of the formal system supporting this description are not given here (see [55]), it is nonetheless possible to make two remarks:

- Heuristics are defined procedurally: roughly speaking they are 'hard-coded', even if the underlying language is declarative;
- A major issue is the management of the multiple heuristics concerning: their idiosyncrasy (who builds them?), their covering range (what psychological domain is covered or not?), their scheduling (how to choose when several heuristics are eligible for execution at the same time?), *etc.*

These criticisms (again, also applicable to the studies of other authors listed above) prompted to propose a model that could be used to implement in a generic manner a whole psychological domain, namely the FFM/NEO PI-R taxonomy. This model is presented in the next sections, beginning with the model of rational agent, on which influence operators are operating.

3. Agent Model

3.1. Agent cycle

Using an approach similar to a BDI-agent cycle process [48][58], we will divide the agent rational reasoning process in a five-step cycle:

1. Deliberation As in BDI theory, given an agent $\mathscr{A}$ with an initial set of possibly contradictory 'desires' $\gamma_i \in \mathbb{G}$, during the deliberation step, the agent chooses within its desires a subset of so-called goals γ_i^* that: 1) it intends to achieve, and 2) must be non contradictory.

2. Planning Given the current goal γ^* in γ_i^*, the agent generates a rational plan $\pi^* \in \mathbb{P}$, which has γ^* in its postconditions: $\gamma \in post(\pi^*)$. A plan is a symbolic expression of the plan language $\mathscr{L}_\pi$ defined in Section 3.2. It is composed of subplans $\pi_i \in \mathbb{P}$ and terminal actions $\alpha_i \in \mathbb{A}$.

3. Decision Although the planner proposes a single plan expression π^* in order to achieve the current goal γ^*, this expression often contains alternatives (equivalent subplans or equivalent terminal actions) and options (optional subplans or optional terminal actions) defined in Section 3.2. Hence, plans are subdetermined and the agent has to make choices between alternatives and to decide whether to execute or not options. Other kinds of decisions are required in the scheduling of the plan operators (discussed in Section 3.2). In any case, the result of the decision process is to elicit the next action α^* of π^* to be executed.

4. Execution The agent executes the current action α^*. In most BDI architectures, this step is implicit: it is explicitly given here in order to support the influence of psychology traits over the execution process itself.

5. Evaluation The agent receives, from the application, information about the actual results of the execution of α^*. It is then possible for the agent to evaluate its performance, both in rational and psychological terms.

3.2. Plan language

The plan language $\mathscr{L}_\pi$ defines expressions that are composed of two kinds of entities: sub plans and terminal actions. In turn, terminal actions are defined in terms of application dependent functions. In the following, when there is no ambiguity, we will simply refer to sub plans as plans, to terminal actions as actions, and to application dependent functions as functions.

3.2.1. Plans

Plans are expressions of the form:

$$\pi = \langle p_i, \gamma_i, pre, post, attr, body \rangle$$

where:
– p_i is a unique plan identifier. Below, we refer to a plan either as π or p_i.
– *pre* is the set of rational preconditions of the plan (noted $pre(p_i)$), *i.e.* facts that *must* hold in the world for π to be applicable.
– *post* is the set of rational postconditions of the plan (noted $post(p_i)$), *i.e.* facts that *necessarily* or *possibly* hold in the world after the execution of π (*cf.* Section 3.2.4).
– $\gamma_i \subseteq post(p_i)$ is the set of goals *of the plan*. It is important to note that the agent may not have the intention to make all the postconditions hold after plan execution, but only the γ_i ones (*cf.* the strategic bomber scenario of Bratman [9]).
– *attr* is the set of attributes defining the plan with regard to psychological influences.
– *body* implements the plan in terms of sub plans and terminal actions (*cf.* Section 3.2.5).

3.2.2. Actions

Actions are expressions of the form:

$$\alpha = \langle a_i, \gamma_i, pre, post, attr, call \rangle$$

where:
– a_i is a unique action identifier.
– *call* is a call to a function $f_i \in \mathbb{F}$, executable in the world.
Other components are defined as in plans above.

3.2.3. Functions

Functions are application dependent basic operations. They are defined in a part of the model dedicated to the application, through an API (namely $\mathbb{F}$) of the form:

$$\varphi = \langle f_i, args, params \rangle$$

where:
– f_i is a unique function identifier, provided by the API of the application.
– *args* is the set of the description of the rational arguments of function f_i. Rational arguments correspond to the so-called *functional proprieties* of the function.
– *params* is the set of the description of the modal arguments of the function. Modal arguments correspond to the so-called *non functional proprieties* of the function (*e.g.* speed, focus of attention *etc.*). We state that $args \cap params = \emptyset$.

3.2.4. Postconditions

Postconditions are facts that hold in the world after the execution of an action or a plan. They can be divided in two main classes:
– Necessary postconditions (noted $post_N$) are sure to hold. For example, given an action a_0 linked to function throw-a-dice, $upface = i$ with $i \in [1..6]$ is a member of set $post_N(a_0) \subseteq post(a_0)$;
– Possible postconditions ($post_P$) are not sure to hold but they could. For example, given a_0, $upface = 6$ is a member of set $post_P(a_0) \subseteq post(a_0)$.

3.2.5. Plan body

The body of a plan has the following syntax:

$$\begin{array}{ll} \text{Body} & \rightarrow \text{BodyProc} \mid \text{BodyDecl} \mid a_i \\ \text{BodyProc} & \rightarrow \text{ProcOp}[\text{Body}|p_i|a_i, \ldots] \\ \text{BodyDecl} & \rightarrow \langle S_p, S_o, S_d \rangle \\ S_p|S_o|S_d & \rightarrow \{p_i|a_i, \ldots\} \end{array}$$

where:
– BodyProc is composed with procedural operators (ProcOp) as defined in Table 4 (given in infix notation).
– BodyDecl is a declarative expression composed of three subsets of subplans or actions: S_p is the set of preferred elements; S_o is the set of optionally executable elements; S_d is the set of default elements.

Table 4
Procedural Operators by descending order of precedence.

Name	Op	Semantics (informal)
seq	;	$a_1;a_2$: Done(a_1) is a precondition to start a_2
alt	‡	$a_1 \ddagger a_2$: only one element is randomly executed
par	$\|\|$	$(a_1 \| a_2) \equiv (a_1;a_2) \ddagger (a_2;a_1)$ one element of the sequence is randomly chosen and executed
case	$\mapsto$	guard$_1 \mapsto a_1$, guard$_2 \mapsto a_2$ guard$_i$ are explicit preconditions for a_i to be executed If several guards are True, one guard is randomly chosen and executed

Plans are synthesized by a rational planner, thus entailing two consequences upon the semantics of BodyDecl expressions:

First, all elements in $\langle S_p, S_o, S_d \rangle$ are not simultaneously executable, but the planner has to make sure that there is always at least one executable element at a given moment (otherwise the agent would enter into a deadlock, which is considered as an error of the rational planner). Since S_p, S_o and S_d are sets, their elements are considered by the rational planner as equivalent alternatives (within a set), at least *wrt* to the goals associated with the body, if not *wrt* to all postconditions.

Second, the meaning attached to 'preferred', 'optional' and 'default' is restricted to the domain of rational planning and is not related to psychology (*e.g.* for 'preferred' see footnote[3]). Thus, a preferred element is meant to generally make the plan more effective, an optional element is meant to be possibly useful for the plan in the future, and a default element is provided by the planner to ensure a minimal effectiveness (*e.g.* to avoid a deadlock issue).

3.2.6. Example

To illustrate the concepts introduced so far, we can consider the following simple plan p, where *pre*, *post* and *attr* sets omitted for concision:

$$p = a_1; a_2 \ddagger a_3; \langle \{g_1\}, \{a_1; a_2, \langle \{g_2, g_3, g_4\}, \{a_1, a_2\}, \\ Null, \{a_7\} \rangle \}, \{a_4\}, \{a_5, a_6, a_7\} \rangle$$
$$\pi_{ex} = < p_{ex}, g_{ex}, \ldots, \{a_1; a_2 \ddagger a_3; < S_{pex}, S_{oex}, S_{dex} >\} >$$

$p \;\; = p_1; p_2; p_3$ // p_1, p_2, p_3 are executed as a sequence
$p_1 \;\; = a_1$ // a subplan can be a terminal action
$p_2 \;\; = a_2 \ddagger a_3$ // random choice between elements a_2, a_3

$p_3 = \langle\{g_1\},\ \{p_4,p_5\},\ \{a_4\},\ \{a_5,a_6,a_7\}\rangle$ // Declarative
$p_4 = a_1;a_2$
$p_5 = \langle\{g_2,g_3,g_4\},\ \{a_1,a_2\},\ Null,\ \{a_7\}\rangle$ // Declarative

Where a Declarative is of the form: $\langle\gamma_i,S_p,S_o,S_d\rangle$.
Note that plan p can also be expressed in short notation by discarding sub plan identifiers:

$p = a_1;a_2 \ddagger a_3;\langle\{g_1\},\{a_1;a_2,\langle\{g_2,g_3,g_4\},\{a_1,a_2\}, Null,\{a_7\}\rangle\},\{a_4\},\{a_5,a_6,a_7\}\rangle$

4. Definition of influence operators

4.1. Definition

An influence operator $\omega \in \mathbb{I}$ is of the form:

$$\omega = \langle o_i, rules, level\rangle$$

Where:
– o_i is a unique operator identifier of the form C_{name}, C being a letter associated with the *class* of the operator (*cf.* below).
– *rules* is a set of (meta) rules adding control over the five steps of the rational process cycle of the agent (*cf.* Section 3.1). They define the operational semantics of the operators as operations on: goals, actions and plans.
– *level* is the level of activation of the operator, which defines how the operator must be applied over the cycle in order to implement a particular trait of personality (see Table 8).

A level is composed of:
– A force $v \in \{0,1,2\}$, used to prioritize the operator (0 means not activated, 1 is activated with low priority (when several operators apply) and 2 is activated with high priority). If different from 0, the value is given as an extra parameter of intensity to the *rules* set (*cf.* class 5 in 4.2).
– A sign $+/-$ (no sign given means +) indicates if the operator implements the concept or its antonym.

4.2. Eight classes of operators

Influence operators can be applied to four kinds of operands: goals (γ), actions (α), plans (π) or functions (φ), depending on their class (*cf.* Table 5). When an operator can be applied to more than one kind of operand, we will use the notation $x|y$ with $(x,y) \in \{\gamma,\alpha,\pi,\varphi\}^2$ and $x \neq y$ (*e.g.* $\alpha|\pi$ means the operand can be an action or plan). We distinguish eight main classes of operators described in the following subsections.

Table 5
Influence operators' operands and step of application.

Class	Operand	Step of application
Goal	γ	Deliberation (1)
Preference	$\alpha\|\pi$	Decision (3)
Filter	$\alpha\|\pi\|\gamma$	Deliberation or Decision (1 or 3)
Scheduling	π	Execution (4)
Modal	$\varphi\|\alpha\|\pi$	Execution (4)
Option	α	Execution (4)
E**X**pectation	$\alpha\|\pi\|\gamma$	Evaluation (5)
Appraisal	$\alpha\|\pi\|\gamma$	Evaluation (5)

4.2.1. Goal operators

Goal operators (G) are used during Deliberation step to control the management of goals, *e.g.* to elicit goals from desires, to arrange goals in the goal stack *etc.*

4.2.2. Preference operators

Preference operators (P) are used during Decision step, to compare and sort in total or partial order a set of rationally equivalent $\alpha|\pi$ (*e.g.* to sort S_p, S_o or S_d sets). Then, it becomes possible to implement mental attitudes such as likes and dislikes, in terms of minimal/maximal elements. In order to compare two items according to a given criterion, one needs a normalized measure function (or cost function) $\chi_{criterion} : \mathbb{F} \mapsto [0.,1.] \cup \{\perp\}$, which implements the criterion. The cost function is application dependent, meaning it has to be defined for each application; if not, it is not an error, it just maps to $\perp$ and is considered void. Hence to each preference operator $P_{criterion}$ is uniquely associated a cost function $\chi_{criterion}$. Comparing functions makes it straightforward to compare actions. By combining action-comparisons, it is often possible to compare plans[3].
For example, let three actions a_1, a_2 and a_3 be respectively linked to functions go-car, go-cycle and go-walk, and the application provides: $\chi_{easy}(\text{go-car}) = 1.$, $\chi_{easy}(\text{go-cycle}) = 0.5$, $\chi_{easy}(\text{go-walk}) = 0.1$. Then a 'lazy' agent could have a preference operator P_{easy} with a high level of activation ($v = 2$), making ac-

[3] In all cases, comparisons are ultimately based on application-dependent information, meaning that two distinct agents elicit the same maximal/minimal elements. Hence our use of the term 'preference' is different from studies based on profiles (*e.g.* preferring red cars over blue cars). We do not consider profile-based preferences operators because they are traditionally integrated into the Planning step (as static external constraints) and they are not considered by psychologists working on personality traits.

tion a_1 to be the maximal element, whenever preferences are required (*e.g.* suppose $S_p = \langle a_1, a_2, a_3 \rangle$ then a_1 will be the psychologically preferred action in a set of rationally preferred actions — respectively for S_o or S_d sets). Note that preference-based operators have to deal with partial orders, and that maximal or minimal elements are not necessarily unique.

4.2.3. *Filter operators*

Filter operators (F) are used to decide whether to keep or to discard an element $\gamma|\alpha|\pi$ for further considering in Deliberation or Decision steps. Filters define the *intrinsic* mental attitude of the agent towards γ, $\alpha|\pi$ (*i.e.* contrary to preferences, no comparison to other elements is performed). For example, if the level of filter operator F_{lawful} is > 0, then action a_1 linked to function kill-people will be discarded, even if it is executable and it was proposed as an efficient solution by the Planning step. Filters have precedence over preferences, *i.e.* preference operators decide among already intrinsically filtered elements.

4.2.4. *Scheduling operators*

Scheduling operators (S) are used during Execution step to control the strategies of plan scheduling. They are divided into heuristic and alteration operators:
– Heuristic operators provide meta control over seq, alt, par, case operators in procedural plans, and over S_p, S_o or S_d sets in declarative plans. For example, consider scheduling operator $S_{predictable}$, which intuitive meaning could be defined as "tending to like executing always in the same order actions that do not require any particular order". If the level of this operator is > 0 then $p_0 = par[a_1, a_2, a_3]$ will always be executed as $seq[a_1, a_2, a_3]$ (Note: executing a par as a seq is rationally correct).
– Alteration operators can alter the structure of procedural plan *e.g.* executing a seq as if it were a par is often (but not always) incorrect; it can be used in case of a very disorderly agent. They can also alter the scheduling process itself *e.g.* by interchanging the position of goals in the goal-stack, or even deleting them. Alteration operators can lead to non rational behaviors, therefore they are not discussed further.

4.2.5. *Modal operators*

Modal[4] operators (M) are used during Execution step to control the actual execution of the function linked to an action, in terms of close related notions of gradualness and intensity, which are provided by application functions. Modal operators can be mapped on plans, recursively down to actions.
– Intensity operators have an influence on functions with intensity *params*. For example, let a_2 be linked to function $\phi = \langle f_2, args, \{speed, \dots\} \rangle$. If the level of the modal operator M_{quick} is $v = 2$ then action a_2 will be executed with maximum value for parameter $speed$. Because $args \cap params = \emptyset$, there is a formal separation between factors relevant to the rational reasoning (mainly the planning process) and to the psychological-based factors; the separation is application-dependent.
– Gradual operators have an influence on functions handling gradual levels of achievement. Several cases of graduality can be considered. The simplest case is a function that can be achieved in a typical manner (like doing the 'par' in golf – not to be mixed with the operator par above), more than the par, less than the par. For example, function give-beggar-money would be affected by the modal operator $M_{nowaste}$, leading it to be either: a) achieved in a typical manner (the 'par' being around 1\$ bill for $M_{nowaste} = 0$); b) overdone (give a 100\$ bill if < -1); c) underdone (give a 5 cents coin if > 1). Note that one has to give at least a non null amount of money otherwise the action would be considered as non-executed.

4.2.6. *Option operators*

Option operators (O) are used during Execution step to add the execution of extra α after (respectively prior to) the execution of the current action. So they are prologue and epilogue add-ons that can modify the postconditions but must not alter the goals (respectively must not alter the preconditions for epilogues). For example, let action a_1 be linked to function open-door and a_2 to pass-door, making the plan $p_1 = a_1; a_2$; moreover, let action a_3 be linked to close-door. If the level of modal operator $O_{cleanup}$ is > 1, then action a_3 can be added as an epilogue to plan p_1. Prologue actions (like wash-hands before doing something) or epilogue actions are application dependent; again if not provided, they are just considered as void influences.

4.2.7. *Expectation operators*

Expectation operators (X) produce expected facts during Deliberation and Decision steps in order to be further used during Evaluation step. They represent the attitude of the agent towards the expected outcomes of an intended γ in terms of its $\alpha|\pi$ postconditions. Three main kinds of mental attitudes can be defined: hope, fear and unconcern. For simplification, we take

[4] Their name comes from the fact they influence modalities (*i.e.* manners) of execution, and do not relate to modal logic.

Table 6
The four types of agent's mental states

	Arity	
Dynamicity	Intrinsic	Interpersonal
Static traits	Features Ψ_F	Relations Ψ_R
Dynamic feelings	Moods Φ_m	Affects Φ_a

here the psychologically-strong position that an agent cannot have attitudes about necessary postconditions $post_N$ (*e.g.* like a stoicist, an agent cannot fear a necessary bad outcome). However agents can have attitudes about possible postconditions. Its attitudes partition the $post_P$ set of $\alpha|\pi$ into three subsets: hoped $post_H$, feared $post_F$ and unconcerned $post_U$, leading to the partition $post = post_N \cup post_H \cup post_F \cup post_U$. For example, taking again action a_0 linked to cast-dice, proposition $upface = 6$ can be both a goal and a hope ($upface = 6 \in post_H$ of a_0).

4.2.8. *Appraisal operators*

Appraisal operators (A) are used during Evaluation step to control the impact upon the mental states of the agent of the actual outcomes of the latest executed item. The general idea is to verify if facts produced by expectation operators actually hold as expected after the execution. For each predicted fact, if the fact indeed holds, there are possible positive effects on mind states (*e.g.* reinforce confidence in one's reasoning capacities) and also possible negative effects (*e.g.* when fears are proven to be confirmed); symmetric effects apply if the predicted fact does not hold.

5. Trait implementation

5.1. *Mental model of the agent*

Typically, the mental states of a person can be classified according to two main criteria:
– *Dynamicity* While several levels of dynamicity of mental phenomena are possible, we will just consider two levels: static traits and dynamic feelings;
– *Arity* Traits and feelings can either be intrinsic to the agent viewed as an autonomous entity, or related to interpersonal relationships, the agent being considered as part of a community of agents.

These two criteria entail a generic classification of mental states in four types, as summarized in Table 6.

Features (Ψ_F) correspond to typical personality attributes that can be considered stable during the agent 'lifetime'.

Relations (Ψ_R) represent generic attitudes of the agent towards other actors in a given community. Note that relations must not be confused with *roles* that define social relations between agents (boss, father *etc.*). Roles have also an influence on rational behaviors but they are external to the personality model. In fact, some Ψ_R describe attitudes of agents towards roles (*e.g.* attitude *wrt* authority). In this paper, we focus on autonomous behavior rather than social behavior so roles will not be discussed further.

Moods (Φ_m) represent intrinsic mental states of the agent that can evolve under the influence of a) external events in the world, b) evaluation of its own actions upon the world.

Affects (Φ_a) represent interpersonal attitudes of the agent that can evolve along its interaction with other agents. While relations are generic, affects are always associated with specific agents in the community.

Despite simplifications, this classification covers most significant notions discussed in mental states modeling [45].

5.2. *Definition of influence operators associated with FFM trait* Conscientiousness

As mentioned in Section 2.3, we rely on a resource where WordNet glosses and Goldberg's q-items are categorized in bipolar behavioral schemes positioned in the FFM/NEO PI-R taxonomy, thus resulting in an enriched taxonomy FFM/NEO PI-R/BS. An excerpt of FFM/NEO PI-R/BS, illustrated in Table 2, shows the two bipolar schemes associated with the first facet of trait Conscientiousness. The total amount of schemes elicited for FFM/NEO PI-R/BS is 69 *i.e.* 138 bipolar positions. Scheme distribution over the FFM/NEO PI-R/BS domain is quite even, with an average of 2.3 schemes per facet. It is interesting to note that the distribution of schemes over the intrinsic/interpersonal trait criterion is also quite even (Ψ_F = 46% and Ψ_R = 54%). Although Ψ_R is certainly worth considering, in this paper we focus on Ψ_F and let Ψ_R for further work.

Considering glosses/q-items contained in the schemes definition resource file, it is possible to exhibit influence operators associated with the described behaviors. In this section, we take as an example only one

Table 8

Activation matrix for the 30 influence operators supporting 12 schemes of FFM factor Conscientiousness.

Facets		Competence		Orderliness		Self-Discipline			Achievement -striving		Deliberation		Dutifulness part.	Ψ_R schemes (not treated)			
Schemes	Generic agent	ASSURED INSECURE ⊕ ⊖	DECISIVE HESITANT ⊕ ⊖	METHODICAL DISORGANIZED ⊕ ⊖	TIDY MESSY ⊕ ⊖	CONCERNED UNCONCERNED ⊕ ⊖	ATTENTIVE INATTENTIVE ⊕ ⊖	MINDFUL FORGETFUL ⊕ ⊖	AMBITIOUS UNAMBITIOUS ⊕ ⊖	HARDWORKER LAZY ⊕ ⊖	STRATEGIC TACTICAL ⊕ ⊖	COMMONSENSE UNSOUND ⊕ ⊖	LAWABIDING LAWLESS ⊕ ⊖	LOYAL UNFAITHFUL ⊕ ⊖	VIRTUOUS IMMORAL ⊕ ⊖	TRUTHFUL UNTRUTHFUL ⊕ ⊖	FAIR PREJUDICED ⊕ ⊖
Operators																	
G_{forget}	1							0 2		0 2				-	-	-	-
G_{quick}	1		2 -2							1 -2				-	-	-	-
G_{deduct}	1										2 0			-	-	-	-
$G_{prudent}$	1										2 0			-	-	-	-
$G_{nowaste}$	1										2 0			-	-	-	-
$G_{knowledge}$	1											2 0		-	-	-	-
G_{reason}	1											2 0		-	-	-	-
P_{easy}	1									0 -2				-	-	-	-
P_{safe}	1	0 2												-	-	-	-
P_{clean}	1				2 -1									-	-	-	-
F_{lawful}	1												2 -1	-	-	-	-
F_{safe}	1						2 0							-	-	-	-
$F_{ambitious}$	1								2 0					-	-	-	-
S_{order}	1			2 -1									2 0	-	-	-	-
$S_{predictable}$	1			2 -1										-	-	-	-
M_{focus}	1	2 -1	1 0	2 -1		2 0	2 0		2 0	2 -2				-	-	-	-
$M_{precision}$	1		1 0			2 0	2 0		2 0	2 -2				-	-	-	-
$M_{tenacity}$	1	2 -2	2 -1			2 0			2 -2	2 -1				-	-	-	-
M_{quick}	1		1 -2							1 -2				-	-	-	-
M_{dopar}	1			-2 2					0 2	2 -2				-	-	-	-
$M_{nowaste}$	1			2 0										-	-	-	-
M_{clean}	1				2 -1					1 -2			2 0	-	-	-	-
O_{extra}	1			0 2						2 0				-	-	-	-
O_{play}	1									0 0				-	-	-	-
$O_{cleanup}$	1				2 0					2 0			2 -1	-	-	-	-
$X_{success}$	1	2 -2							2 0					-	-	-	-
X_{hope}	1	2 -2							2 0					-	-	-	-
$X_{choiceConf}$	1	2 -2	2 -2											-	-	-	-
$A_{responsible}$	1					2 0				2 -1				-	-	-	-
$A_{success}$	1								2 - 2	1 -1				-	-	-	-

FFM trait: Conscientiousness, because it is dealing mostly with Ψ_F. It is composed of 6 NEO PI-R facets containing 16 schemes, as given in Table 8. In trait Conscientiousness, 11 schemes have influence operators $\in \Psi_F$, while 4 others $\in \Psi_R$ (LOYAL, VIRTUOUS, FAIR, TRUTHFUL) are not treated. Scheme LAWABIDING, having influence operators in both sets, is partially treated. A set of 30 influence operators required to support the 12 Ψ_F-related schemes of FFM trait Conscientiousness (called $\mathbb{I}_C$) are listed in Table 7, with their informal semantics given as a gloss.

Examples of operator elicitation Considering in turn each of the 69 schemes in FFM/NEO PI-R/BS, it is possible to interpret its glosses in terms of the eight classes of influences over the BDI cycle.

As a first example, one can consider the glosses and q-items of scheme TIDY/MESSY in Table 2:

	pole+ TIDY
597	clean or organized
1035	marked by order and cleanliness in appearance or habits
Q71	Like order
Q72	Like to tidy up
Q73	Want everything to be 'just right'
Q74	Love order and regularity
	pole- MESSY
375	lacking neatness or order
918	dirty and disorderly
Q76	Often forget to put things back in their proper place
Q77	Leave a mess in my room
Q78	Leave my belongings around
Q79	Am not bothered by messy people

Table 7

$\mathbb{I}_C$: operator set for Conscientiousness

Op. name	Gloss of + pole (informal semantics)
$G_{forget}+$	forgets γ (*e.g.* drops from goal stack)
G_{quick}	deliberates quickly (*e.g.* takes first in list of alternatives)
$G_{deduct}+$	controls the depth of deduction process
$G_{prudent}$	considers consequences of γ
$G_{nowaste}$	dislikes γ that waste resources
$G_{knowledge}+$	* has a large knowledge base (facts)
$G_{reason}+$	* has a large common sense base (heuristics)
P_{easy}	prefers easy $\gamma\|\pi$
P_{safe}	prefers safe $\gamma\|\pi$
P_{clean}	prefers clean $\gamma\|\pi$
F_{lawful}	discards unlawful $\gamma\|\alpha\|\pi$
F_{safe}	discards unsafe $\gamma\|\alpha\|\pi$
$F_{ambitious}$	prefer ambitious $\gamma\|\alpha\|\pi$
S_{order}	executes $\alpha\|\pi$ in order (- pole: seq⇒par)
$S_{predictable}$	executes $\alpha\|\pi$ routinely (*e.g.* par⇒seq)
M_{focus}	executes $\alpha\|\pi$ with mental concentration
$M_{precision}$	executes $\alpha\|\pi$ with precision
$M_{tenacity}$	perseveres if failures/difficulties during $\alpha\|\pi$
M_{quick}	executes $\alpha\|\pi$ quickly
M_{dopar}	control graduality of execution (see Section 4.2 #5)
$M_{nowaste}$	don't waste resources while executing $\alpha\|\pi$
M_{clean}	is clean while executing $\alpha\|\pi$
$O_{extra}+$	controls general tendency to execute optional α from S_o
O_{play}	adds pleasurable α, taken from S_o
$O_{cleanup}$	adds clean-up action(s) (not in S_o) after execution of $\alpha\|\pi$
$X_{success}$	expects success of current $\gamma\|\alpha\|\pi$
X_{hope}	has high level of hopes in $poss(\alpha\|\pi)$
$X_{choiceConf}$	confident in its choices of $\gamma\|\alpha\|\pi$
$A_{responsible}$	feels responsible for the results of its $\gamma\|\alpha\|\pi$
$A_{success}$	has a particular sensitivity to success

* operator overlaps with capacities in rational reasoning.

\+ negative level values don't apply for these operators.

Given on the one hand the eight classes of influences proposed in the model (and summarized in Table 7), and on the other hand the glosses and q-items of scheme TIDY/MESSY, one has to establish possible correspondences between them. Elicited relations are annotator dependent, hence ideally involve a multi-annotator process in several phases (not developed in this paper). In the particular case of the TIDY/MESSY example, typically one would be able to associate its glosses and q-items with three different classes (*P*, *M* and *O*), and in turn, relate this scheme to three operators:

Operator	Informal semantics
P_{clean}	Preference for clean entities *e.g.* actions, objects, people
M_{clean}	Execute an action in a clean manner, if available
$O_{cleanup}$	Put the world into place after execution *e.g.* close doors

As a second example, one can consider glosses and q-items of scheme ASSURED/INSECURE = {lacking self-confidence, showing fear…} (*cf.* Table 2 for more glosses), which suggest a preference for safe $\alpha|\pi$ (therefore associated to operators $P_{safe}|F_{safe}$), a lack of focus and tenacity in execution (corresponding to $M_{focus}|M_{tenacity}$), *etc.*

5.3. Definition of activation matrix

Given a set of operators elicited from a trait, it is then possible to define an implementation of its schemes in terms of activation levels (as defined in Section 3.2) of these influence operators. Then, the levels of activation control the cycle of the rational agent defined in Section 3.1, according to their class, defined in Section 4. Activation levels are defined formally with two main principles:

1. They are relative values, describing a positive/negative tendency/deviation from an average behavior supposed to describe a 'normal' person, further called the *generic agent*;
2. They are defined in a so-called *activation matrix* of schemes × operators, where each cell contains by default the activation level of the generic agent except in cases where the glosses/q-items of the scheme specify a deviation from average behavior.

Actual activation levels must be elicited by annotators, which have to be persons with expertise in psychology. The description of the complete controlled process, with several expert annotators involved in a multi phase discussion, is beyond this presentation since it is too extensive and mainly related to psychology issues. Hence example developed below, we just present an instance of activation matrix (*cf.* Table 8) as it has resulted from such a process.

5.4. Generic and specific agent

The formal definition of the notion of *generic agent* is still an open issue from a psychological perspective and it exceeds this study. In the general case that could be carried out in further works, one can consider grounding the notion of generic agent through the use of statistical data about a human population; then the generic agent would correspond to the mean behaviors associated with traits of the members of that population. However, in the restricted context of computational studies, we are not concerned with the a definite definition; rather we focus on the mechanisms that implement a given definition onto the decision making process of a rational agent. Hence one can use a first order approximation of the notion of generic agent, which can be based on an ontology built by experts: such an ontology has been elicited in previous works [8] and presented in short in Section 2.3.

Consequently, we will take here a restricted position where a generic agent is defined in terms of low levels of activation for the influence operators. A strict definition $\forall \omega_i, level(\omega_i) = 0$ is not correct because most people have some degree of confidence, orderliness, dutifulness *etc.* corresponding to a positive, low/medium activation level *i.e.* 1 (except for some traits like Neuroticism where it would be -1). When several operators are to be applied in the same cycle step and they have the same level of activation, they must be prioritized to avoid contradictory influences. For a generic agent, a static pre-order on schemes can be predefined.

5.5. *Designing a specific character*

To keep the example simple, we consider here a generic agent having a low level of activation for each operator in Table 7: $\forall \omega_i \in \mathbb{I}_C$, $level(\omega_i) = 1$. The activation matrix of 12 Ψ_F-related schemes of Conscientiousness is given in Table 8 where empty values = 1 and non empty values describe the deviation from generic agent's values.

Given Table 8, one can define a specific character (*i.e.* an artificial agent exhibiting a specific set of psychological phenomena) by selecting which schemes it is likely to exhibit. Typically, a person only exhibits few tendencies deviating from average behavior; hence in a human-like agent, few schemes will be selected (say two or three). Note that formally, it is possible to select an arbitrary number of schemes. When more than one scheme is selected, it can happen that for a given operator ω_i, two schemes induce conflicting levels (*e.g.* a TIDY and LAZY agent for M_{clean}). Similarly to the generic agent, it is possible to predefine a static pre-order on schemes. Moreover, when human-like agents are defined, conflict of operators can be a source of realism (*e.g.* by making obvious the agent's hesitation between two conflicting tendencies such as TIDY and LAZY).

6. Conclusion and future works

In this paper, we have presented an approach based on the principle that personality traits a) can influence the deliberation cycle of rational agents and b) can be defined in terms of level of activation of influence operators. Using our R&B framework, we have shown that it is possible to tweak the deliberation cycle of BDI agents, and thus described a set of eight classes of influence operators. Operators from those eight classes have been used in the particular case of one of the five main classes of the FFM taxonomy, Conscientiousness, resulting in a set of 30 operators.

In this study, we have focused on intrinsic personality traits (Ψ_F), which cover only 46% of the behavioral schemes. Remaining schemes are related to interpersonal relationships (Ψ_R), which require a deliberation cycle adapted to interaction. Moreover we have stated in Section 1.2.3 four main restrictions that we plan to address in further work. Especially we intend to extend our model to handle Ψ_R-based traits, following the same approach, and ultimately, to propose a full set of operators that would allow experimenters to implement any combination of facets from the FFM/NEO PI-R taxonomy. Another development will be the study of characters containing complex characters entailing conflicting influences over the rational process. Finally, it will also be important to consider validating how human users perceive the personality of agents implemented using the R&B framework on some specific scenarios.

References

[1] J. F. Allen, H. A. Kautz, R. N. Pelavin, and J. D. Tenenberg. *Reasoning About Plans*. Representation And Reasoning. Morgan Kaufmann Publishers, Inc., 1991.

[2] G. W. Allport. *Personality: A psychological interpretation*. New York: Holt, 1937.

[3] J. R. Anderson. ACT: a simple theory of complex cognition. *American Psychologist*, 51(4):355–365, 1996.

[4] J. Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7):122–125, 1994.

[5] Y. Bellik, I. Rebai, E. Machrouh, Y. Barzaj, C. Jacquet, G. Pruvost, and J. P. Sansonnet. Multimodal interaction within ambient environments: an exploratory study. In *Porceedings of INTERACT 2009*, volume 1, pages 89–92. IFIP, LNCS, Uppsala, Sweden, 2009.

[6] F. Bouchet and J. P. Sansonnet. Caractérisation de requêtes d'Assistance à partir de corpus. In Lang J., Y. Lespérance, D. Sadek, and N. Maudet, editors, *Actes des Quatrièmes Journées Francophones Modèles Formels de l'Interaction (MFI'07)*, volume 8 of *Annales du LAMSADE*, pages 269–276, Paris, France, 2007. Université Paris Dauphine.

[7] F. Bouchet and J. P. Sansonnet. A framework for modeling the relationships between the rational and behavioral reactions of assisting conversational agents. In *Proc. of the 7th European Workshop on Multi-Agent Systems (EUMAS'09)*, Agia Napa, Cyprus, 2009.

[8] F. Bouchet and J. P. Sansonnet. Classification of wordnet personality adjectives in the NEO PI-R taxonomy. In *Fourth Workshop on Animated Conversational Agents, WACA 2010*, pages 83–90, Lille, France, 2010.

[9] M. E. Bratman. *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[10] M. E. Bratman. What is intention? In P. R. Cohen and M. E. Morgan, J.and Pollack, editors, *Intentions in Communication*, pages 15–32. The MIT Press, Cambridge, MA, 1990.

[11] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Villiamsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *CHI'99: Proceedings of the SIGCHI conf. on Human factors in comp. syst.*, pages 520–527, New York, NY, USA, 1999. ACM Press.

[12] R. B. Cattell, H. W. Eber, and M. M. Tatsuoka. *Handbook for the sixteen personality factor questionnaire (16 PF)*. Champain, Illinois, 1970.

[13] P. T. Costa and R. R. McCrae. *The NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, 1992.

[14] M. Dastani. 2APL: a practical agent programming language. In *AAMAS '08:The seventh international joint conference on Autonomous agents and multiagent systems*, volume 16, pages 214–248, Estoril, Portugal, 2008. Springer-Verlag.

[15] M. Dastani and J. J. Meyer. Programming emotional agents. *International Journal of Intelligent Systems*, 25(7):636–654, 2010.

[16] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989.

[17] A. Dillon and M. G. Morris. User acceptance of new information technology: theories and models. *Annual Review of Information Science and Technology*, 14(4), 1996.

[18] T. Doce, J. Dias, R. Prada, and A. Paiva. Creating individual agents through personality traits. In *Intelligent Virtual Agents (IVA 2010)*, volume 6356 of *LNAI*, pages 257–264, Philadelphia, PA, 2010. Springer-Verlag.

[19] A. Doniec, R. Mandiau, S. Piechowiak, and S. Espie. Controlling non-normative behaviors by anticipation for autonomous agents. *Web Intelligence and Agent Systems*, 6(1):29–42, 2008.

[20] P. C. Ellsworth and K. R. Scherer. Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, and H.H. Goldsmith, editors, *Handbook of affective sciences*, Series in affective science, pages 572–595. Oxford University Press, New York, NY, USA, 2003.

[21] R. Evertsz, F. E. Ritter, P. Busetta, and M. Pedrotti. Realistic behaviour variation in a BDI-based cognitive architecture. In *Proc. of SimTecT'08*, Melbourne, Australia, 2008.

[22] S. B. G. Eysenck, H. J. Eysenck, and P. Barret. A revised version of the psychoticism scale. *Journal of Personality and Individual Differences*, 6(1):21–29, 1985.

[23] C. Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[24] N. H. Frijda. *The Laws of Emotion*. Psychology Press, 2006.

[25] J. A. Garcia-Pardo and C. Carrascosa. Social welfare for automatic innovation. In F. Klugl and S. Ossowski, editors, *Multiagent System Technologies*, volume 6973 of *Lecture Notes in Computer Science*, pages 29–40. Springer Berlin / Heidelberg, 2011.

[26] L. R. Goldberg. Language and individual differences: The search for universal in personality lexicons. *Review of personality and social psychology*, 2:141–165, 1981.

[27] L. R. Goldberg. An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology*, 59:1216–1229, 1990.

[28] J. Gratch and S. Marsella. A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*, 5(4):269–306, 2004.

[29] B. Hayes-Roth, L. Brownston, and R. van Gent. Multiagent collaboration in directed improvisation. In *Proc. of the First Int. Conf. on Multi-Agent Systems ICMAS 95*, pages 148–154, San Francisco, CA, 1995.

[30] N. Howden, R. Rannquist, A. Hodgson, and A. Lucas. Intelligent agents - summary of an agent infrastructure. In *Proc. of the 5th International Conference on Autonomous Agents*, Montreal, 2001.

[31] H. Jiang, J. M. Vidal, and M. N. Huhns. eBDI: an architecture for emotional agents. In *AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–3, New York, NY, USA, 2007. ACM.

[32] O. P. John, R. W. Robins, and L. A. Pervin, editors. *Handbook of Personality: Theory and Research*. The Guilford Press, 3rd edition, 2008.

[33] J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: an architecture for general intelligence. *Artif. Intell.*, 33(1):1–64, 1987.

[34] M. Lim, J. Dias, R. Aylett, and A. Paiva. Improving adaptiveness in autonomous characters. In H. Prendinger, J. Lester, and M. Ishizuka, editors, *Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, pages 348–355. Springer, 2008.

[35] P. Maes. Agents that reduce work and information overload. *Commun. ACM*, 37(7):30–40, 1994.

[36] P. Maes. Artificial life meets entertainment: lifelike autonomous agents. *Commun. ACM*, 38(11):108–114, 1995.

[37] L. Malatesta, G. Caridakis, A. Raouzaiou, and K. Karpouzis. Agent personality traits in virtual environments based on appraisal theory predictions. In *Artificial and Ambient Intelligence, Language, Speech and Gesture for Expressive Characters,achie AISB'07*, Newcastle, UK, 2007.

[38] D. P. McAdams. A conceptual history of personality psychology. In R. Hogan, J. A Johnson, and S. R. Briggs, editors, *Handbook of personality psychology*, pages 3–39, San Diego, CA, US, 1997. Academic Press.

[39] R. R. McCrae and P. T. Costa. Validation of the five-factor model of personality across instruments and observers. *J. of Personality and Social Psychology*, 52(1):81–90, 1987.

[40] M. McRorie, I. Sneddon, E. de Sevin, E. Bevacqua, and C. Pelachaud. A model of personality and emotional traits. In *Intelligent Virtual Agents (IVA 2009)*, volume 5773 of *LNAI*, pages 27–33, Amsterdam, NL, 2009. Springer-Verlag.

[41] X. Min, L. Padgham, A. Mbala, and J. Harland. Tracking reliability and helpfulness in agent interactions. *Web Intelligence and Agent Systems*, 5(1):31–46, 2007.

[42] W. Mischel, Y. Shoda, and O. Ayduk. *Introduction to Personality*. John Wiley & Sons, Inc., eighth edition, 2008.

[43] H. J. Mosler. Modelling social behavior with a socio-psychological simulation approach. *Web Intelligence and*

Agent Systems, 2(3):185–200, 2004.

[44] J. Nielsen. *Usability Engineering*. Academic Press, Boston, 1993.

[45] A. Ortony. On making believable emotional agents believable. In R. Trappl, P. Petta, and S. Payr, editors, *Emotions in humans and artifacts*, pages 189–211. MIT Press, Cambridge, MA, 2003.

[46] A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge, UK, cambridge university press edition, 1988.

[47] C. Pelachaud. Some considerations about embodied agents. In *Int. Conf. on Autonomous Agents*, Barcelona, 2000.

[48] A. S. Rao and M. P. Georgeff. BDI agents: From theory to practice. In *Proc. First Int. Conference on Multi-agent Systems (ICMAS-95)*, pages 312–319, 1995.

[49] P. Rizzo, M. V. Veloso, M. Miceli, and A. Cesta. Personality-driven social behaviors in believable agents. In *AAAI Symposium on Socially Intelligent Agents*, pages 109–114, 1997.

[50] D. Rousseau. Personality in computer characters. In *AAAI Technical Report WS-96-03*, pages 38–43, 1996.

[51] N. Sabouret and J. P. Sansonnet. Learning collective behavior from local interactions. In *CEEMAS: Second International Workshop of Central and Eastern Europe on Multi-Agent Systems*, volume 2296 of *LNCS*, pages 273–282, Cracovia, Poland, 2002. Springer Verlag.

[52] N. Sabouret and J. P. Sansonnet. Un langage de description de composants actifs pour le Web sémantique. *Information - Interaction - Intelligence Review, Revue I3*, 3(1):9–36, 2003.

[53] J. P. Sansonnet and F. Bouchet. Expression of behaviors in assistant agents as influences on rational execution of plans. In *Intelligent Virtual Agents (IVA 2010)*, volume 6356 of *LNAI*, pages 413–419, Philadelphia, PA, 2010. Springer-Verlag.

[54] J. P. Sansonnet and F. Bouchet. Extraction of agent psychological behaviors from glosses of wordnet personality adjectives. In *Proc. of the 8th European Workshop on Multi-Agent Systems (EUMAS'10)*, Paris, France, 2010.

[55] J. P. Sansonnet and F. Bouchet. Joint handling of rational and behavioral reactions in assistant conversational agents. In Helder Coehlo, Rudi Studer, and Michael Wooldridge, editors, *Proc. of the 19th European Conference on Artificial Intelligence (ECAI 2010)*, pages 1049–1050, Lisbon, Portugal, 2010. IOS Press.

[56] J. P. Sansonnet and F. Bouchet. Integrating psychological behaviors in the rational process of conversational assistant agents. In R. C. Murray and P. M. McCarthy, editors, *Proceedings of the 24th International FLAIRS Conference (FLAIRS-24)*, pages 440–445, Palm Beach, Florida, 2011.

[57] J. P. Sansonnet, D. Leray, and J. C. Martin. Architecture of a framework for generic assisting conversational agents. In *International Conference on Intelligent Virtual Agents IVA'06*, volume 4133 of *LNAI*, pages 145–156, CA: Marina del Rey, 2006. Springer.

[58] S. Sardina. A course on BDI architectures. `http://goanna.cs.rmit.edu.au/~ssardina/courses/Roma07PhDcourse/`, 2007.

[59] B. G. Silverman, M. Cornwell, and K. O'Brien. Human behavior models for agents in simulators and games. In *PRESENCE*, volume 15, pages 139–162, 2006.

[60] B. Steunebrink, M. Dastani, and J. J. Meyer. A logic of emotions for intelligent agents. In R. C. Holte and A. E. Howe, editors, *Proceedings of AAAI-07*, pages 142–147. Vancouver, Canada, 2007.

[61] T. van der Weide, F. Dignum, J. J. Meyer, H. Prakken, and G. Vreeswijk. Personality-based practical reasoning. In I. Rahwan and P. Moraitis, editors, *Argumentation in Multi-Agent Systems*, volume 5384 of *LNCS*, pages 3–18. Springer, 2009.