Intelligent Agents with Personality: from Adjectives to Behavioral Schemes

François Bouchet and Jean-Paul Sansonnet
McGill University, Canada and LIMSI-CNRS, France

ABSTRACT

Conversational agents are a promising interface between humans and computers; but to be acceptable as the virtual humans they pretend to be, they need to be given one of the key elements used to define human beings: a personality. As existing personality taxonomies have been defined only for description, we present in this chapter a methodology dedicated to the definition of a computationally-oriented taxonomy, in order to use it to implement personality traits in conversational agents. First, a significant set of personality-traits adjectives is registered from thesaurus sources. Then, the lexical semantics related to personality-traits is extracted while using the WordNet database and it is given a formal representation in terms of so-called Behavioral Schemes. Finally, we propose a framework for the implementation of those schemes as influence operators controlling the decision process and the plan/action scheduling of a rational agent.

INTRODUCTION

Context: Intelligent Agents for Assisting Human Users

Intelligent agents are autonomous software entities which perceive their environment and acts on it to accomplish their goals. Among elements of the environment of such agents are human beings. Some categories of agents have been more and more in interaction with human users, sometimes becoming themselves the interface between the users and the system they want to use. Those are called the *conversational agents* (Maes, 1994; Cassell, Bickmore, Billinghurst, Campbell, Chang, Villiamsson & Yan, 1999), and as many tend to be there to provide them with some kind of *assistance*, we'll therefore focus here on this subcategory of agents that are the *intelligent assistant agents* (further referred to simply as *agents*). In that kind of situation, three entities are in bilateral interaction (a human user U, an intelligent assistant agent A and a computer system S), where the user performs some activity on/with the system, and, at times, they can solicit the agent for general advice or for direct help upon the system or the task at hand, in which case the agent might have to interact with the system directly, on behalf of the user. In such a situation (further called a UAS situation), one can see that agents need to be both able to interact with:

- a symbolic model of the application and a rational reasoning capacity about that model. In the following, we will refer to this part of an agent as the "*rational agen*t". Actually, intelligent assistant agents should be able to interact with a system in the same way as autonomous agents achieve practical reasoning, in order to perform tasks in a given environment.
- the user, which requires: a) a conversational interface, which is often multimodal; b) processing, in a rational way, the input of user's requests in order to produce as output factual replies c) expressing the answers in a way that would look and sound similar to how a human being assisting the user in the same task would.

Personality Traits for Agents: Why and How?

That last point requires taking into account what we know of the cognitive processes of human beings, in order to mimic it in a software agent. It means that it is not enough for an agent to be *technically competent* thanks to their symbolic reasoning capacities: they also need to be *psychologically relevant,* to be acceptable (in the sense of (Davis, 1989)), especially when they deal with people of the general public. One of the key points to ensure the consistency of an agent in the way it will provide information to the users is for it to be given what makes a human being follow the same usual patterns of behavior and emotions: *personality traits*. Having agents with personality traits is also particularly interesting in order to try to provide different kind of assistance strategies, which could be adapted to the personalities of the human users.

Those considerations led us to propose a framework dedicated to the study of the nature of the relationships between the rational and the psychological reasoning capacities: the Rational and Behavioral architecture (R&B), where "behavioral" has the particular meaning of "psychological behavior", and which can be found at http://www.limsi.fr/~jps/research/rnb/rnb.htm. R&B is a generic framework enabling the computational definition and the experimentation of various rational/psychological strategies. In recent work based on R&B (Sansonnet & Bouchet, 2010), we have proposed a model where psychological phenomena, such as personality traits, are implemented in terms of influence operators altering the rational process of an agent. Presently, the proposed model has been implemented only with few arbitrary-chosen psychological phenomena (e.g. a *lazy*/*gloomy*/*cheerful*… agent) (Bouchet & Sansonnet, 2009). This approach illustrates the principle and the feasibility of a framework like R&B, but it fails to cover significantly the main domain of the description of the psychology of a person, especially in terms of personality traits. Therefore, there is a real need for a more systematic and computationally tractable definition of the domain, covering most psychological phenomena. To do that, computer scientists can start from the literature on psychology, which provides various taxonomies of personality traits. Unfortunately, those are rarely fine-grained enough (despite the existence for some of them of second level elements called facets, like (Costa & McCrae, 1992)) and they are too generic to allow a straightforward computational implementation.

Concretely, we propose in this chapter to follow a methodology that can be summarized in the following four key-points:

1. We suggest refining the facets in order to propose a three-level taxonomy for describing the personality of a person (and ultimately, of an agent).

2. The third level of this extended taxonomy will be composed of so-called behavioral schemes that are intended to provide implementable instantiations of the facets, especially in terms of influences upon actions and plans of the rational agent.

3. We propose to elicit those schemes from a corpus of personality adjectives, which are in turn associated with their lexical senses and glosses in the WordNet lexical database (Fellbaum, 1998). Not only linguistics resources have long been a source for building psychological, but WordNet also offers three extra advantages: it is a computerized database, it has a standard position in natural language processing, and, above all, it provides a comprehensive set of lexical semantics senses (called *synsets*) that will provide a solid ground for defining the schemes.

4. The scheme elicitation process is conducted in two main phases:
   - Phase 1: Selection of a significant subset of personality adjectives and classification into a standard, two-level taxonomy (traits-facets)
   - Phase 2: For each facet, clusterization of involved lexical senses into schemes.

In order to better understand our proposition, it appears necessary to provide readers that might not be very familiar with this field some background information about agents in general, and psychology research on personality, which we will do in a first background section. In a second section, we will develop the aforementioned methodology inn four points. Finally, we'll conclude and review further works that the existence of our extended taxonomy and its implementation within an agent framework make possible.

BACKGROUND

First, we will focus on the definition of rational agents and provide an overview of the existing frameworks in this domain. Then we will introduce the major works of past decade that tried to add some personality elements to the traditional rational agents, and their motivation to do so. In a third subsection, we will then focus more particularly on the class of conversational agents (differing from rational ones by the fact they are in dialogical interaction with humans), and review attempts to bring personality elements into those. Finally, in a fourth subsection, we will review research in psychology to see, independently from the agent perspective, how researchers in that field have been treating the concept of (human) personality, which methodologies were used and to what kind of taxonomies they led.

Frameworks for Rational Agents

In our context, the term *"rational agent"* will be compliant with traditional definitions issuing from Artificial Intelligence (AI) and Multi-agents Systems (MAS). For example, Russell & Norvig (2009) state that "*a rational agent is an AI software component that does the right thing*" while achieving commonsense reasoning, using its own current knowledge upon the state of the world (pp. 38-39). Wooldridge (2000) provides a similar definition for MAS agents by stating that "*an agent is said to be rational if it chooses to perform actions that are in its own best interests, given the beliefs it has about the world*" (p.1). According these traditional definitions, rational agents are associated with programs capable of *practical reasoning*, i.e. building plans and choosing actions to be executed in order to achieve their goals.

One of the first attempts at modeling the cognitive reasoning process of an agent is the outstanding SOAR (State, Operator And Result) architecture (Laird, Newell & Rosenbloom, 1987; Anderson, Matessa, & Lebiere, 1997) which uses explicit IF-THEN rules and can synthesize plans according to several strategies (problem space search, means-ends analysis, hill-climbing etc.). Another noticeable architecture is ACT-R (Adaptive Control of Thought-Rational) (Anderson, 1996), which operates on similar principles and focuses on the separation between declarative and procedural representations (actually, the list of existing cognitive architectures is large: CHREST, CLARION, ICARUS, DUAL, PSI…).

Following SOAR or ACT-R architectures, the BDI (Beliefs, Desires, Intentions) approach developed by Rao & Georgeff (1995) is a theory of practical reasoning (deciding what to do next cf. (Bratman, Israel, & Pollack, 1988)) directed towards situated reasoning about actions and plans (Allen, Kautz, Pelavin, & Tenenberg, 1991). BDI architectures recognize the primacy of *beliefs*, *desires* and *intentions* in rational action.
It is common for BDI platforms to use practical reasoning rules (PR-rules) of the form: < Head→ Guard: Body >, where the Head is associated to a goal, the guard to a belief (often expressed as a predicate function over the world) and the body to a procedure (an instantiated plan). The process of triggering the proper rule at a given moment is called the *deliberation cycle*.
Many operational frameworks implement the BDI model such as JAM (Huber, 1999), JACK (Howden, Rönnquist, Hodgson, & Lucas, 2001), 2APL (Dastani, 2008), etc.
These architectures have been quite successful at creating both autonomous and multi-agent systems capable of operating in computational contexts, and are therefore good candidates for the support of intelligent assistant agents.

A generic version of the deliberation cycle of a BDI architecture as given by authors such as (Wooldridge, 2000) or (Sardina, 2007) is as follows:

Algorithm 1: Agent Control Loop
while true **do**
1 observe the world;
2 update internal world model; // assess the current state of affairs
3 deliberate about what intention to achieve next; // what to do next
4 use means-ends reasoning to get a plan for the intention; // how to do it
5 execute the plan; // do it
end

Steps 3 and 4 are specific to the BDI theory:
- Given an agent A with an initial set of possibly contradictory desires g_i, during step 3, the agent chooses within its desires a subset of so-called goals g* such that: 1) A intends to achieve goals in g* and 2) goals in g* must be non-contradictory (whereas desires g_i can be contradictory).
- Traditional rational reasoning is then performed (step 4) to synthetize all the plans, and within a plan all the sub-plans that will enable the agent to reach a goal in the set g*. Indeed, although the planner proposes a single plan expression in order to achieve the current goal, this expression often contains alternatives (equivalent sub-plans or equivalent terminal actions) and options (optional sub-plans or optional terminal actions). Hence, plans are sub-determined 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. In any case, the result of the decision process is to elicit the next atomic action to be executed.

Note that practical reasoning (deciding what to do next) is often opposed to *theoretical reasoning* (deciding what is True/False), which is directed towards the logical consistency of the knowledge base of the agent and will not be addressed here. Moreover, we also discard from our definition the Decision Theory approach, where an agent is said rational if it maximizes a function of expected utility (Russell & Norvig, 2009). For that, we rely on the argument of those authors stating that contrary to practical reasoning, *decision theory* is based on total knowledge (actually, decision theory is much addressed in current multi-agent system studies in social and economics macro-models).

Adding Personality Traits to Rational Agents

Recently, authors have attempted to integrate into rational agent architectures some psychological notions, giving two main reasons to do so: first, to propose more complete cognitive models of agents; second, to propose agents capable of sustaining more human-like interactions with people, especially ordinary people involved in conversational activities with intelligent assistant agents.

While some authors have proposed a model of emotions based on traditional SOAR architecture (Gratch & Marsella, 2004), with a significant impact upon the inner SOAR architecture, most authors have preferred to propose improvements of BDI architectures because their deliberation cycle is more flexible. In this case, they exhibit both rational reasoning modules and psychological reasoning modules (Dias & Paiva, 2005; Lim, Dias, Aylett, & Paiva, 2008). In these architectures, the psychology of the agent influences both the expression of emotions through the body modalities (the agents are embodied as graphical virtual characters for example) but also has an impact on the decision making process. For example, the eBDI model implements emotions in a BDI framework (Jiang, Vidal, & Huhns, 2007). They give a good introduction about the history of the necessity to implement emotions into rational agents: “According to Weiss (1999) there are four types of traditional agent architectures: logic based agents;

reactive agents; layered architectures; BDI agents", but according to Jiang *et al.*, all these approaches have drawbacks:

- In logic based agents, decision making is predicated on the assumption of calculative rationality: the world will not change in any significant way while the agent is deciding what to do, but most current multi-agent systems cannot guarantee a static and deterministic environment.
- Reactive agents make decisions based on local information, so they must have sufficient information available in their local environment for them to determine an acceptable action, and it is difficult to see how such decision making could take into account non-local information.
- Layered architectures are very general. The main problem is that while they are arguably a pragmatic solution, they lack the conceptual and semantic clarity of un-layered approaches. Another issue is that of interactions among layers.
- BDI architectures are known to reflect human practical reasoning process, and they have both widely accepted philosophical and logical roots. However, Jiang et al. claim that BDI architectures currently ignore the influence of emotions in decision-making.

Using the well-used agent creation platform JACK (Howden, Rennquist, Hodgson & Lucas, 2001) that implements the BDI theory, CoJACK (Norling & Ritter, 2004; Evertsz, Ritter, Busetta, & Pedrotti, 2008) is an extension layer intended to simulate physiological human constraints like a) the duration taken for cognition; b) working memory limitations (e.g. "losing a belief" if the activation is low or "forgetting the next step" of a procedure); c) fuzzy retrieval of beliefs; d) limited focus of attention or e) the use of moderators to alter cognition. One can note that the cognitive notions handled in CoJACK are not closely related with the psychology traits of a human person, but rather to generic mental capacities. Nonetheless, CoJACK remains an outstanding attempt at introducing cognitive notions into formal frameworks for rational agents.

Adding Personality Traits to Conversational Agents

SOAR, ACT-R and BDI frameworks have been primarily developed to support *artificial* intelligent agents: in a single-agent context for SOAR or ACT-R, and in single/multi-agent context for BDI. In these systems, an agent interacts with the environment or with other artificial agents but not directly with human beings. On the contrary conversational agents are primarily designed to support extensive multimodal conversational interacting sessions with people, often persons of the general public. This is the reason why there has been a much stronger incentive for introducing psychological notions into their architecture, which is also often based on rational agent engines.

Historically, the first works in this area were performed in the context of so-called *"believable agents"*, introduced in the mid-1990s. Several research projects and groups (the Carnegie Mellon University Oz project (Bates, 1994), the Agents group MIT (Maes, 1995), the Virtual Theater project at Stanford (Hayes-Roth, Brownston & van Gent, 1995), etc.) have focused their efforts on the creation of computerized believable agents.

Since then, there have been various attempts at implementing psychological features into conversational cognitive architectures.
The works of Rousseau & Hayes-Roth (1996) stated the principle that personality traits actually express themselves as influences on actions and plans. Another major contribution is the idea that personality traits can be associated with intensity factors so that a computational planning engine can take them into account. Rizzo, Veloso, Miceli, & Cesta (1997) prove 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), and once chosen, the goals are planned and executed. Malatesta, Caridakis, Raouzaiou, & Karpouzis (2007) use traits to create different expressions of behaviors, especially by influencing the

appraisal part of the OCC theory (Ortony, Clore, & Collins, 1988; Ortony, 2003). Their work focuses on how agents evaluate 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. The PMFserv framework (Silverman, Johns, Cornwell, & O'Brien, 2006) is dedicated to the creation of culturally credible agents by using performance moderators functions (PMFs). The authors claim that "*its principal feature is a model of decision-making based on emotional subjective utility constrained by stress and physiology*". This work was partly based on Gillis & Hursh's,(1999) work, introducing the notion of *behavior moderators*. Like CoJACK, they focus on a differentiation based on the physical capacities of the agents, as their aim is to simulate crowd behavior in military forces.
All these authors emphasize the fact that agents should exhibit both a rational reasoning engine and a psychological reasoning engine.

Recently, Allbeck & Kress-Gazit (2010) have proposed a framework where complex instructions can be given by the user, especially by using natural language to control an agent. In that case, the agent is a robot (it is an autonomous agent but not strictly speaking a conversational agent) acting in a complex environment. The user's instructions constrain the execution of the agent's plan. This work is focused on dynamic actions and planning for robots. The constraints applied to the actions and plans are provided by an external source (a human user), and can sometimes impact over the action and plans (e.g. prohibiting an action, changing the actions execution order, etc.).
In the same way, works of McRorie, Sneddon, Sevin, Bevacqua, & Pelachaud (2009), based on the well-used architecture of conversational agent GRETA (Pelachaud, 2000), involve models of personality for the expression of emotions (face, gesture etc.). And as claimed by Doce, Dias, Prada, & Paiva (2010), it is necessary to create individual agents through personality traits because "*in the era of globalization, concepts such as individualization and personalization become more and more important in virtual systems"*. They personally also use the Five Factor Model (Goldberg, 1992) as a basis for creating different personalities, by using traits that can automatically influence cognitive processes such as appraisal, planning or coping. They also consider there is a need for an emotion model that can easily represent emotions in a systematic way.

Works on Personality Traits Classification

Although presented above as separate notions, the rational and the psychological reasoning capacities of actually work in quite an intricate manner (Ellsworth & Scherer, 2003; Frijda, 2006). Moreover, many studies mentioned above focus on low-level/transitory psychological notions such as emotions and moods, but other notions like personality traits (John, Robins, & Pervin, 2008) that are associated with high-level/long-lasting features of the personality of a human being, should be integrated and could be promising for developing agents with consistent characters.

Authors in philosophy and in psychology have attempted to define the notion of personality, such as:

- An individual's characteristic pattern of thinking/feeling/acting;
- A person's general style of interacting with the world;
- Differences between people which are relatively consistent over time and place

According to Neill (2008), there have been four main historical periods, corresponding to four main paradigms: Psychoanalysis (1900's—); Behaviorism and Classifications (1950's—); Humanistic (1960's—); Social-Cognitive Transpersonal (1980's—):

1. *Psychoanalytic* by Freud is a psychodynamic perspective (there is a strong notion of personality development) which emphasizes on unconscious motivation. It is intended for therapy but it is also a theory of personality involving three main concepts: a) the *id* is mainly biological and deals with aggression and pleasure. It is unconstrained by logic or reality and does not distinguish between reality and fantasy; b) the *ego* is about rationality; and c) the *superego* is dealing with societal and moral values. It strives for the ideal and is responsible for guilt.

2. *Behaviorism and Classifications* is mainly based on personality traits. A trait is traditionally defined as: a characteristic pattern of behavior or a disposition to feel and act. In this domain, two main approaches are prevailing: a) *type-based classifications* label each person as a single type/class; b) *traits-based classifications* identify the degree to which several different personality factors occurring within an individual (notion of independent dimensions). Nowadays, the four main Personality Inventories are: EPQ (Eysenk & Eysenk, 1975); 16PF (Cattell, Eber, & Tatsuoka, 1970); MMPI (Graham, 2006) and FFM (Goldberg, 1990).
3. *Humanistic* by Maslow (1968), along with Rogers, which were discontent both with Freud's negativity and the mechanistic psychology of the behaviorists and trait-based classifications. Their approach a) emphasizes on healthy individuals and defense maneuvers (like coping) and b) defines the well-known *scale of needs*: Physiological/Safety/Belonging/Esteem/Self-actualization.
4. *Social-cognitive* by Bandura (1986) combines social learning and cognition. They claim that human behaviors emerge from the interplay between a person and her/his environment (according to a principle of reciprocal determinism). This is an optimistic view that allows growth and change based on person centered therapy.

In the following, we will rely upon the approach based on behaviorism and classifications, because it is the most comprehensive and tractable from the computational point of view. Moreover, we will focus on trait-based classifications because there is a large consensus about their usefulness at least from the standardization point of view. When one is interested in the taxonomy of the psychological phenomena, particularly those related to personality traits, the two most successful paradigms are:

- *The two dimensional model of Eysenck* (EPQ), which is mainly based on personality questionnaires, which are questionnaires (generally with yes/no questions) to assess the personality traits of a person (Eysenk & Eysenk, 1975; Eysenck, Eysenck & Barret, 1985);
- *The Five Factor Model* (FFM), which is mainly based on lexical data. It is a convergent research from many authors in psychology, from Norman (1963) to Goldberg (1990), etc.

Following early work of Cattell's personality trait classification into 16 factors (Cattell, Eber & Tasuoka, 1970), which was still prominent in the 1980s, Eysenck developed in 1976 a model based on a biological approach, mainly stating that personality traits are heritable and have a foundation in the nervous system (Eysenck, 1976). In this view, they are supposed to be very stable parameters of human psychology. The three basic traits found by Eysenck were: Psychotism, Extraversion, Neuroticism (PEN), with a high and a low pole. Later, Eysenck focused on a representation of personality traits based on two axes only: Neuroticism (also denoted *emotional stability*) and Extraversion which entails four cases: phlegmatic, melancholic, choleric and sanguine, encompassing 32 basic facets. Eysenck then proposed a set of criteria as an attempt to merge the three major systems, namely Cattell's 16 PF, Norman's Five Factor Model, and Eysenck's PEN system (Eysenck & Eysenck, 1975).

The FFM approach to the taxonomy of traits is based on natural language and more precisely lexical resources (Baumgarten, 1933), such as the glosses found in dictionaries. The *lexical hypothesis* states that most of socially relevant and salient personality characteristics have become encoded in the natural language (Allport, 1937). The lexical approach has been promoted by Goldberg who claimed that "*personality vocabulary provides an extensive, yet finite, set of attributes that people speaking a given language have found important and useful in their daily interactions*" (Goldberg, 1981). In 1990, Goldberg tried to define a small set of 475 common trait adjectives grouped into 131 sets of factors that can cover the big five domain (Goldberg, 1990). It issued in 1992 into the 50-item instrument using so-called 'transparent format' (Goldberg, 1992) that finally led to the modern version of FFM. The FFM is based on five large classes of psychological traits, often named Big Five model or OCEAN, by taking the first letter of the name of each class (the classes are listed in the first column of Table 1).

Table 1: Facets for enriching the FFM classes: three approaches (alignment according to John et al. *judgment, based on (Soto & John, 2009)). Facet names in italics correspond to a secondary possible attachment of a facet.*

OCEAN Classes	Lexical facets (N=18) (Saucier & Ostendorf, 1999)	NEO PI-R facets (N=30) (Costa & McCrae, 1992)	CPI-Big five facets (N=16) (Soto & John, 2009)
Extraversion	Sociability	-	-
	Assertiveness	Gregariousness	Gregariousness
	-	Activity	-
	Activity/Adventurousness	Excitement-seeking	*(O Adventurousness)*
	-	-	Social confidence (vs Anxiety)
	-	Positive emotions	-
	Unrestraint	-	-
	(A Warmth/Affection)	Warmth	-
Agreeableness	Warmth/Affection	*(E Warmth)*	-
	Modesty/Humility	Modesty	Modesty (vs Narcissism)
	-	Trust	Trust (vs Suspicion)
	-	Tender-Mindedness	Empathy/sympathy
	Generosity	-	Altruism
	Gentleness	Compliance	-
	-	Straightforwardness	-
Conscientiousness	Orderliness	Order	Orderliness
	Industriousness	Achievement striving	-
	Reliability	Dutifulness	Industriousness
	Decisiveness	-	-
	-	Self-Discipline	Self-discipline
	(O Perceptiveness)	Competence	-
	-	Deliberation	-
Neuroticism	Insecurity	Anxiety	Anxiety
	Emotionality	-	-
	Irritability	Angry hostility	Irritability
	-	Depression	Depression
	-	-	Rumination-compulsiveness
	-	Self-consciousness	*(E Social confidence (vs Anxiety))*
	-	Vulnerability	-
	-	Impulsiveness	-
Openness	Intellect	Ideas	Intellectualism
	-	Aesthetics	Idealism
	Imagination/creativity	Fantasy	-
	-	-	Adventurousness
	-	Actions	-
	-	Feelings	-
	-	Values	-
	Perceptiveness	-	-

Note that while it is largely used today, the FFM model has undergone many critics since its beginning. For example, McAdams (1992) proposed an early critical appraisal of the FFM mode with six main arguments, often developed by other authors afterwards: 1) it is limited to personality *traits*; 2) it lacks actual behavior prediction power; 3) it does not address causal explanations of human behavior; 4) it does not consider any representation or knowledge about the world; 5) it does not provide a program for human personality integration and improving; 6) it relies on the notion of s*hort phrases* (taken either from dictionaries or questionnaires) for the description of the psychology of a person. McAdams claims that FFM describes a *generic* personality for humankind not the personality of this man (that I know of) or this woman (that I know of) by stating: "*the Five-factor model is essentially a psychology of a stranger*".

The FFM being a very generic classification, several authors have tried to refine this taxonomy by dividing the FFM classes into so-called facets (Saucier & Ostendorf, 1999; Costa & McCrae, 1992; Soto & John, 2008). John, Naumann & Soto (2008a) have shown that these facet lists have many similarities

although their facet number varies from 16 in (Saucier & Ostendorf, 1999) to 30 in the so-called NEO PI-R (NEO PI-R stands for: **N**euroticism **E**xtraversion **O**penness **P**ersonality **I**nventory **R**evisited) proposition of Costa and McCrae (1992). The 30 facets of the NEO PI-R classification are listed in the second column of Table 1, together with their glosses. NEO PI-R is a long standing model that provides a very precise facet list. For these reasons, we will rely on it for the work presented in the next sections.

BUILDING A RESOURCE FOR PERSONALITY DESCRIPTION

In this section, we will describe our contribution to the aforementioned domains by defining a computationally-oriented resource of personality for conversational agents. In a first subsection, we shall justify our choice of a lexical approach of personality and in particular the use of WordNet as a resource. In a second subsection, we will explain the methodology followed to classify personality adjectives within the NEO PI-R taxonomy. Then we will describe how that work allowed eliciting some concrete human behavior schemes associated to the taxonomy, based on the adjective glosses. Finally, in a fourth section, we will show how those schemes can lead to a concrete implementation within a particular agent framework as influence operators.

Lexical Semantics Senses of Personality Adjectives

Following the lexical hypothesis, we are interested in the sub set of the vocabulary that describes the personality of a person (further called personality adjectives) in order to assess the actual behaviors of people that are associated with such adjectives. Actually we are not interested in the adjectives, as words, but rather in the lexical semantic senses (in short the senses) of adjectives that describe the personality of people, in terms of their actual behaviors. Such instances of human behaviors, related to personality adjectives can be exhibited from so-called *short phrases* or *glosses*. There are three main sources for short phrases: a) Dictionary senses definitions; b) Personality questionnaires; c) Electronic lexical databases.

As an illustration of what dictionary senses definitions look like, here is the entry of the Merriam-Webster (MW) dictionary for personality adjective 'lazy' (which might not be the most important trait we would like to integrate into an agent, but provides a rich example here):

LAZY:
1 a: disinclined to activity or exertion: not energetic or vigorous
1 b: encouraging inactivity or indolence <a lazy summer day>
2: moving slowly: sluggish
3: droopy, lax <a rabbit with lazy ears>
4: placed on its side <lazy E livestock brand>
5: not rigorous or strict <lazy scholarship>

This entry contains five main distinct senses, the first one being divided into two sub cases. For most senses, one or several short phrases are given as instances of the corresponding sense. They are the behavioral source we are looking for in order to carry a computational implementation of the sense. Note that not all of the senses are related to personality description, e.g. droopy, lax <a rabbit with lazy ears> refers to physical proprieties of objects or people.

An alternative source to elicit the personality traits of a particular person is provided by the questions (in this context, called *items*), which are hand-built for the constitution of so-called personality questionnaires. Below is an excerpt of 10 first yes/no questions from the revised version of the original 100 items of the EPQ Model of personality of Eysenck & Eysenck (1975).

Occupation: ... Age: ... Sex: ...

Instructions: Please answer each question by putting a circle around the 'YES' or the 'NO' following the question. There are no right or wrong answers, and no trick questions. Work quickly and do not think too long about the exact meaning of the questions.

1. Do you have many different hobbies?	YES	NO
2. Do you stop to think things over before doing anything?	YES	NO
3. Does your mood often go up and down?	YES	NO
4. Have you ever taken the praise for something you knew someone else had already done?	YES	NO
5. Do you take much notice of what people think?	YES	NO
6. Are you a talkative person?	YES	NO
7. Would being in debt worry you?	YES	NO
8. Do you ever feel 'just miserable' for no reason?	YES	NO
9. Do you give money to charities?	YES	NO
10. Were you ever greedy by helping yourself to more than you have share of anything?	YES	NO

… and 90 more.

Because an item like “Are you a talkative person?” is linked with a specific personality trait (say E-Gregariousness), in an outside table, a simple transposition of the questionnaire makes it possible, for a given trait, to list the set of its associated behaviors.

Previously, we have mentioned that authors already have classified sets of adjectives in the FFM model and in the NEO PI-R model. However, when we tried to use these classifications in our works, we encountered two main problems:
- *Access to data*: Classification tables provided in the papers are too synthetic to be exploited and it is not easy to access the original resource data files.
- *Computational approach to semantics*: In order to fulfill the requirements of computational definition, we need words but also the precise lexical semantics attached to them (their precise senses). For example, the word ‘kind’ has three senses: /Tolerant/, /Genial/, /Openhearted/ (see Table 2) that can be classified into two distinct FFM classes and three distinct NEO PI-R facets: resp. A-compliance, E-warmth, A-tendermindedness. This example shows that the traditional word-level is insufficient. Instead we need to work at the sub level of the semantic senses attached to words.

The WordNet lexical database (Fellbaum, 1998) comes handy when one has to treat a large amount of lexical data and one has to access the lexical semantics of words. In WordNet, a word is attached to several so-called *synsets* that define a unique lexical sense described by a gloss. Moreover, because the WordNet database is freely accessible, it makes it easy to build a computer aided system for the classification process. Below, in order to compare with traditional dictionaries like MW, is the WordNet definition for personality adjective 'lazy', which has two senses/synsets (noted here between / /):

LAZY:

1. SYNSET /Workshy/	disinclined to work and exertion	= sense 1.a	in MW
2. SYNSET /Slow/	moving slowly and gently	= sense 2	in MW

One can see that there is some correspondence (but not bijective) between the Merriam-Webster definition and the WordNet definition for personality adjective 'lazy'.

*Table 2: WordNet synsets associated with three commonly used personality adjectives. In column two, synsets with * marks are not related to personality traits.*

Adjective	Not	Synset	WordNet gloss defining the synset
friendly		/Pally/	characteristic of or befitting a friend
	*	/Allied/	of or belonging to your own country's forces or those of an ally
	*	/Easy/	easy to understand or use
		/Favorable/	inclined to help or support; not antagonistic or hostile
kind		/Tolerant/	tolerant and forgiving under provocation
		/Genial/	agreeable, conducive to comfort
		/Openhearted/	having or showing a tender and considerate and helpful nature; used especially of persons and their behavior
lively		/Vital/	full of spirit
	*	/Eventful/	filled with events or activity
		/Frothy/	full of life and energy
	*	/Springy/	elastic; rebounds readily
		/Alert/	quick and energetic
		/Racy/	full of zest or vigor

Classification of a Set of Personality Adjectives into the FFM/NEO PI-R Taxonomy

In previous work (Bouchet & Sansonnet, 2010), we have carried out a classification of a set of personality adjectives in the FFM/NEO PI-R taxonomy. Below, we give a summary of this process carried out by two expert annotators, in three main steps:

Step 1: Personality adjective selection.

In order to work on actually used personality adjectives, we have collected a set of personality adjectives C_{coll} from ten different Internet sources explicitly claiming to provide "*lists of adjectives describing personality traits*" (the sources are listed in (Bouchet & Sansonnet, 2010)). The set C_{coll} contains 1055 distinct adjectives, providing a first order approximation of the linguistic domain related to personality traits adjectives. Moreover, for each personality adjective a_i in C_{coll}, we have associated a salience rank ra_i depending on the number of lists in which a given adjective appears, which represents its usage frequency in the resources files. We have observed that ra_i values are ranging from 1 to 9. Unfortunately, some of the words appearing in set C_{coll} are not present in the WordNet database (|29|) or are present but are not listed as adjectives (|21|). However, their low salience (ra_i =1) allowed us to choose to discard them, leading to a set C_{WN} of 1005 distinct WordNet-tractable adjectives. Table 3 gives an excerpt of C_{WN}, which shows that the coverage of English language is quite large, if not exhaustive.

Table 3: Excerpt of personality adjectives from the C_{WN} set containing 66 adjectives beginning with letter 'a' (listed not by salience rank but in alphabetic order).

abandoned, abashed, aberrant, abhorrent, abiding, abject, able, abnormal, abrasive, abrupt, absent-minded, absorbed, absurd, abusive, accepting, accomplished, acid, active, adamant, adaptable, addicted, adorable, adventurous, affective, afraid, aggressive, agonizing, agreeable, alcoholic, alert, alive, all-around, alluring, aloof, ambiguous, ambitious, amoral, amused, amusing, anal-retentive, ancient, angry, animated, annoyed, annoying, anxious, apathetic, appealing, appreciated, argumentative, arrogant, articulate, artificial, artistic, ashamed, assertive, astonishing, athletic, attentive, auspicious, automatic, awake, aware, awesome, awful, awkward.

Step 2: Selection of the relevant adjective-synset-gloss triplets.
Each adjective a_i in the set C_{WN} has a WordNet entry, which associates with a_i a set of lexical semantics senses (i.e. its synsets). In the WordNet base, each synset is a unique identifier and defines a unique sense by means of a short phrase called a gloss. It is thus possible to obtain from the C_{WN} set of adjectives, a base B_{all} of triplets of the form: < adjective, synset, gloss >. However as one can see from entries marked with a mark * in Table 2, not all triplets are related to the description of the personality of a person. In order to discard these entries, an annotating process was performed by two expert annotators (details are given in (Bouchet & Sansonnet, 2010)), which resulted in the selection of a base B_{sel} of 1356 relevant triplets, containing 904 distinct adjectives.

Step 3: Classification of the triplets into the FFM/NEO PI-R taxonomy.
The NEO PI-R taxonomy proposes for each of the five FFM personality classes 6 facets, thus resulting in 30 clusters. In NEO PI-R, each facet encompasses both the concept and its antonym(s), e.g. the facet A-Modesty can stand for adjectives like '*mild*' as well as '*arrogant*'. Hence, similarly to Goldberg's works on 50 bipolar scales (Goldberg, 1992), each facet was divided into two poles: noted + for the concept and - for its antonym(s), thus resulting in 60 target clusters noted +/-NEO PI-R. The classification was performed by two expert annotators (details in (Bouchet & Sansonnet, 2010)), which resulted in the building of a base, further transformed into an XML resource. This resource is freely available for research and teaching purposes on the Web [http://perso.limsi.fr/jps/research/rnb/toolkit/taxo-glosses/taxo.htm]. Table 4 displays an excerpt of this XML resource listing triplets associated with FFM trait Openness.

A first attempt at classifying a subset of B_{sel} entries over the 60 established classes raised a major issue: despite the fact that the collected adjectives words were claimed by the resources to be personality-related, actually a lot of their corresponding adjective-synset-gloss triplets in WordNet are also associated with other attributes of a person, which can be divided into five main categories, called *external categories*:

- physical descriptions, e.g. "having a sturdy and well-proportioned body" (athletic, /Muscular/);
- mental capacities e.g. "characterized by quickness and ease in learning" (bright, /Smart/), (smart, /Bright/);
- judgmental qualities, both physical or mental, e.g. "appealing to the emotions as well as the eye" (lovely, /Beautiful/);
- social roles, e.g. "characterized by or indicating poverty" (poor, /Beggarly/);
- mind states are transitory (like emotions), e.g. "in an aroused state", (excited, /Thrillful/).
- errors are items that don't fit in any of the facets of categories.

The final distribution of gloss-oriented triplets into the categories (30 facets + external cases) is summarized by Figure 1. Although some facets like C-Dutifulness and A-Tender-mindedness appear to have gathered significantly more glosses, we can see that all NEO PI-R facets have had some associated triplets (with a minimum of two for O-Aesthetics – see Table 4, line Facet: Aesthetics pole +). Moreover, some negative positions have no triplets attached to them. This is due to the fact that they were introduced systematically as complement to positive poles – for example, see Table 4, line Facet: Aesthetics pole -.

Figure 1: Distribution of B_{sel} triplets over the FFM/NEO PI-R taxonomy

Table 4: Classification of B_{sel} triplets into the FFM/NEO PI-R taxonomy. An excerpt of trait Openness *is given for covering its facets:* Fantasy, Aesthetics, Feelings, Action, Ideas, Values *(facet* Ideas *is abridged; facet* Values *is no listed). Triplets are separated into two sets: pole + (the concept of the facet) and pole – (its antonym). Each line displays a triplet in the format:* gloss [adjective, Synset (with '/' omitted), rank].

FFM trait: Openness
Facet: Fantasy pole + not sensible about practical matters; idealistic and unrealistic [romantic, WildEyed, 5] having the ability to produce or originate [productive, Generative, 4] concerned with sacred matters or religion or the church [religious, Spiritual, 4] promoting construction or creation [creative, Constructive, 3] having the ability or power to create [creative, Originative, 3] marked by independence and creativity in thought or action [imaginative, Imaginative, 3] concerned with or affecting the spirit or soul [spiritual, Unearthly, 3] showing great reverence for god [godly, Worshipful, 2] Facet: Fantasy pole - unimaginatively conventional [conservative, ButtonDown, 3] guided by practical experience and observation rather than theory [practical, Hardheaded, 3] aware or expressing awareness of things as they really are [realistic, Virtual, 3] freed from illusion [disillusioned, Disillusioned, 2] sensible and practical [earthy, DownToEarth, 2] denying or questioning the tenets of especially a religion [skeptical, Unbelieving, 2] concerned with the world or worldly matters [mundane, Terrestrial, 1] deficient in originality or creativity; lacking powers of invention [sterile, Uninspired, 1]
Facet: Aesthetics pole + marked by refinement in taste and manners [cultured, Cultured, 5] relating to or characteristic of art or artists [artistic, Art, 1] Facet: Aesthetics pole -
Facet: Feelings pole + able to feel or perceive [sensitive, Sensible, 7] given to or marked by sentiment or sentimentality [sentimental, Tender, 2] Facet: Feelings pole - tired of the world [bored, WorldWeary, 3]
Facet: Action pole + possibly accepting or permitting [capable, Open, 4] capable of adapting (of becoming or being made suitable) to a particular situation or use [adaptable, Adaptable, 3] willing to undertake or seeking out new and daring enterprises [adventurous, Adventurous, 3] disposed to venture or take risks [daring, Venturous, 3] fearless and daring [bold, Vaulting, 2] Facet: Action pole - excessively fastidious and easily disgusted [nice, Prissy, 6] incapable of adapting or changing to meet circumstances [inflexible, Person, 2]
Facet: Ideas pole + eager to investigate and learn or learn more (sometimes about others' concerns) [curious, Snoopy, 5] having curiosity aroused; eagerly interested in learning more [curious, Interested, 5] ahead of the times [modern, Innovative, 5] open to arguments, ideas, or change [receptive, Pervious, 5] ready or willing to receive favorably [receptive, Open, 5] devoted to matters of the mind [reflective, Intellectual, 5] ready to entertain new ideas [open-minded, Open-minded, 3] having an open mind [hospitable, Open, 2] of high moral or intellectual value; elevated in nature or style [idealistic, Rarified, 2] … Etc.

Elicitation of Behavioral Schemes Associated with Adjectives

From the process of classification of the triplets into the FFM/NEO PI-R taxonomy we have obtained a resource base where each facet contains two bipolar lists of triplets. The glosses contained in a given facet provide several instantiations of the psychological notion associated with the facet.

Actually, these instantiations can describe *distinct human behaviors* needing to be implemented with distinct computational heuristics. For example, considering the facet Openness-Fantasy in Table 5, classified glosses are actually dealing with three distinct notions: creativity, idealness and spirituality, which are mixed in both the + or - pole of the facet.

This is the reason why we have to carry out a more precise clustering of the glosses within the facets, in terms of so-called behavioral schemes (in short *schemes*), where each scheme can reasonably be viewed as "an atomic kind of behavior". Indeed, each scheme is intended to be further associated with a specific computational implementation, for example in terms of operators influencing decisions and choices in the deliberation cycle of the BDI agents of the R&B framework.

Each scheme is a structure of the form:

Scheme ID
FFM/NEO PI-R category
Positive pole ID, Textual definition, {$gloss_1$, .. $gloss_N$}
Negative pole ID, Textual definition, {$gloss_1$, .. $gloss_N$}

Table 5: Clusterization of facet Openness-Fantasy (Ofa) *into three distinct atomic bipolar schemes. For poles +/- of scheme* CREATIVENESS*, we give the set of its associated triplets, actually abridged here to the gloss part (text italicized)*

Scheme's poles identifiers	**Textual definition of the scheme** + abridges 'more than average' resp. 'less than average'	**Gloss count**
1. IDEALISTICNESS	Ofa = Openness-fantasy	
+ IDEALISTIC	+ inclined to like (even to believe in) imaginary ideas or situations	1
- PRACTICAL	- inclined to like (or to believe in) imaginary ideas or situations	5
2. SPIRITUALISTNESS	Ofa	
+ SPIRITUALIST	+ interested in spirit and soul matters	6
- MATERIALIST	- interested in spirit and soul matters, even denying the tenets	2
3. CREATIVENESS	Ofa	
+ CREATIVE	+ having power of creativity in thought or action • *Having the ability to produce or originate* • *Promoting construction or creation* • *Having the ability to create* • […] *marked by independence and creativity in thought or action*	4
- UNIMAGINATIVE	- having power of creativity in thought or action • *Unimaginatively conventional* • *Deficient in originality or creativity; lacking powers of invention* • *Unimaginative and conformist* • *Lacking spontaneity or originality or individuality*	4

Like facets, schemes are organized into a positive pole and a negative pole. Contrary to facet-poles that support a two-position scale (concept/antonym opposition), scheme-poles are organized onto a three-position scale: more than average (+), average (=), less than average (-). The average position implicitly is associated with the "*average behavior associated with the personality of a generic human being*" while +

and - poles explicitly denote a person with a marked psychological inclination. Technically, the semantics of a scheme is only and uniquely defined by the list of its associated glosses. However, each scheme pole is provided with an informal textual definition synthetizing its glosses.

During the process of organization of the facets into behavioral schemes 766 glosses were annotated and clustered into 69 distinct schemes resulting in a total of 2 × 69 = 138 scheme +/- poles.

Figure 2: Distribution in FFM/NEO PI-R taxonomy of glosses (top) and schemes (bottom), where: N_{facet} = 30, N_{gloss} = 766, N_{scheme} = 69, mean number of glosses per facet = 26 and mean number of schemes by facet = 2.3

Figure 2 displays the distribution of the glosses and of the schemes in the FFM/NEO PI-R taxonomy. Globally, one observes a good and homogeneous coverage in terms of glosses and schemes over the five OCEAN traits and over the 30 NEO PI-R facets. There is also a good and homogeneous coverage of glosses over the schemes: the distribution varies from min = 2 to max = 40, with a mean of 26 glosses by scheme.

Association of Influence Operators to the Schemes

Once the WordNet glosses of the set of personality adjectives have been reorganized into the FFM/NEO PI-R facet taxonomy, they must be associated with an actual computational implementation. In the R&B framework, this part is taken by the so-called influence operators (further called *operators*), which are symbolic rule-based heuristics operating on/altering the rational behavior of the agent (see (Bouchet & Sansonnet, 2009) for a formal definition). In this context, a final step towards the computational implementation of the schemes is required because there is often but not always a direct correspondence between a scheme and an operator. It comes from the fact that schemes are concepts elicited from the linguistic domain point of view, whereas the operators we wish to define are symbolic heuristics (or meta rules) defined from a computational point of view. This heterogeneity entails three issues:

- It can happen that a scheme requires several distinct operators to be implemented;
- Conversely, the same operator can implement distinct schemes, meaning that actually these schemes are not distinct from a computational point of view;
- Moreover, the +/- poles of the schemes can be inverted with the +/- poles of the operators (especially in the Neuroticism class).

From the set of 69 schemes, it was possible to elicit a set of 56 operators, which can be organized in a two-level taxonomy. Whereas the two-level scheme taxonomy is psychology-oriented (FFM/NEO PI-R) the two-level operator taxonomy is computation-oriented, that is organized according to the activities of the agent. Because we consider conversational agents in UAS situation, as stated in the introduction, the first level of the classification of the operators separates:
- *Mono-agent situations*: with operators mainly related to the A ↔ S interactions. This class of operators is made of four subsets: personal knowledge; personal action; personal satisfaction; personal control of moods/affects.
- *Multi-agent situations:* with operators mainly related to the U ↔ A interactions. This class of operators is made of three subsets: knowledge exchange; multi-agent action and control of others; interactional feelings and manners.

Table 6 lists the taxonomy of the No = 56 operators and for each category the count of its operators. From this classification we could draw three main results:

Table 6: Two-level taxonomy of the influence operators (No = 56)

1. Mono Agent	$N_{po} = 26$	2. Multi Agent	$N_{io} = 30$
Personal knowledge: Memory, fact, commonsense, learn	4	**Interaction - knowledge**: trust, frankness, curiosity, privacy, conversion	5
Personal execution of action and plans: adaptation, decision, deliberation, creativity, danger, expectation, novelty, clean, method, concern, attention, activity, work	13	**Interaction - action and control of others**: Meeting, cooperation, sacrifice, giveness, fairness, forgiveness, commitment, charisma, domination, menace, agreement, honesty, malice, conform, progress	15
Personal satisfaction: possession, success, pleasure, art	4	**Interaction - feelings and manners**: comfort, reassurance, perception, empathy, humility, tolerance, aggression, friendliness, showiness, smalltalk	10
Personal mood/affect control: moodiness, emotionality, fierceness, assurance, prone(mood)	5	-	

1. First level balance: There is a good balance between the two sets of personal operators ($N_{po} = 26$) and the interactional operators ($N_{io} = 30$). This balance is not artificial: it is the result of the distribution of the lexical semantics attached to the initial set of personality adjectives. Actually, the linguistic senses reflect the two faces of an agent: its personal face and its social face.

2. Scheme-operator alignment: The alignment between schemes and operators is quite good. Actually, most schemes (68%) are in a one to one relation with an operator and 25% of the others operators occur only in two schemes, as shown as follows (where $\text{prone}(\text{mood}_i)$ wraps in a single class the tendency to exhibit 7 kinds of moods):

Type of operators counted	Frequency	% of N_O
All occurrences in schemes	82	-
With 1 occurrence (bijections)	38	67.9%
With 2 occurrences	14	25.0%
With 3 occurrences	2	5.4%
With special operator $\text{prone}(\text{mood}_i)$	7	12.5%

3. Corpus coverage: A third result is the good convergence of personality adjectives over operators: repetitive random samples of only 100 adjectives, taken from C_{WN} (*i.e. samples of* 10% of the 1005 adjectives of C_{WN}) have proved to cover about 90% of the 56 operators. Hence, we can consider that the resource base built in this work provides a good coverage of the personality trait domain in terms of schemes and operators.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

While software agents are more and more in interaction with human users, there is a real need for more human-like interactions between ordinary people and artificial agents. Hence, artificial agents should not only be mere rational entities, providing efficient services but also friendly entities, playing the mediating roles involved in presentation, assistance, teaching, game companions, coaching, etc.
A way to fulfill these requirements is to provide traditional rational agent architectures with specific features dealing with psychological notions such as personality traits and/or moods. However, this can only be achieved: 1) when a solid ground for the description of the personality of a generic person is defined, according to state of the art works in psychology; but also 2) when personality traits descriptions are computationally implementable, which entails a need for behavioral precision.

We have defined and described a taxonomy of psychological behaviors through a study based 1) on standard personality traits taxonomies in psychology (FFM with NEO PI-R for the facets) and 2) on a commonly used lexical database (WordNet). Using the lexical semantic senses of an original selection of 1005 personality adjectives in that base, we have dispatched them over the 60 facets of NEO PI-R, taking into account their polarity. The glosses associated to those senses in WordNet thus provide extra and useful information about how to computationally implement a given facet. We have finally shown that although there is often a direct mapping between a psychological facet and a computational operator, there are several exceptions, which led us to define a second computationally-oriented taxonomy.

We believe that this taxonomy can become a useful resource (currently available on the Internet) for future research works focusing on the implementation of psychological behaviors in terms of the elicited operators. Once implemented, an operator would then have to be evaluated. Considering the class of multi-agent operators, for example in the context of conversational agents, it can be done by making human subjects interact with the agent. Considering the class of mono-agent operators, for example in the context of autonomous agents, it can be done by observing the agent actual actions. These experiments should confirm that people actually would notice differences in behaviors, and that those differences actually correspond to the personality traits or facets that were intended to be implemented by that operator.

From a more application-based point of view, the fact that this taxonomy has been conceived for a concrete computational application (agent-based interfaces) makes it particularly useful for agent framework designers, as opposed to taxonomies originally made exclusively to describe human personality. We have shown that they could be implemented within our own agent framework (R&B) as what we called influence operators: a similar work could be done within other existing agent frameworks, by modifying them to include such operators.
As for system designers who simply wish to add a conversational agent within their system, although choosing how their agent(s) should react in the particular context of their application still remain exclusively their decision, they can at least use this taxonomy as an exhaustive set of personality traits and associated behaviors they should consider during the agent design. Such a process would have to be done in two phases: firstly, by asking themselves what each of those influence operators would mean for them (e.g. “what would it mean for our particular agent to be trustful towards the user?”), and secondly, once this question has been answered, to decide if they want to implement it or not (e.g. “do we want our agent to behave this way or not?”), in an ad hoc way, if they don’t rely on an existing agent framework.

REFERENCES

Allbeck, J. M., & Kress-Gazit, H. (2010). Constraints-Based Complex Behavior in Rich Environments. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent Virtual Agents* (Vol. 6356, pp. 1-14). Berlin, Heidelberg: Springer Berlin Heidelberg.

Allen, J. F., Kautz, H. A., Pelavin, R. N., & Tenenberg, J. D. (1991). *Reasoning about plans*. (R. J. Brachman, Ed.). Morgan Kaufmann Publishers Inc.

Allport, G. W. (1937). *Personality: a psychological interpretation*. Oxford, England: Holt.

Anderson, John R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*(4), 355-365.

Anderson, John Robert, Matessa, M., & Lebiere, C. (1997). ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention. *Human-Computer Interaction*, *12*(4), 439.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Social Learning Theory. Englewood Cliffs, NJ, US: Prentice-Hall.

Bates, J. (1994). The role of emotion in believable agents. *Commun. ACM*, *37*(7), 122-125.

Baumgarten, F. (1933). Die Charaktereigenschaften [The character traits]. (A. Francke, Ed.)*Beiträge zue Charakter-und Persönlichkeitsforschung*, (1).

Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems* (pp. 592–603). Maceio, Alagoas, Brazil: Springer.

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

Bouchet, F., & Sansonnet, J.-P. (2010). Classification of Wordnet Personality Adjectives in the NEO PI-R Taxonomy. *Actes du Quatrième Workshop sur les Agents Conversationnels Animés (WACA 2010)* (pp. 83-90). Presented at the WACA 2010, Lille, France.

Bratman, M. E. (1990). What Is Intention? In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 15-32). The MIT Press.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and Resource-bounded Practical Reasoning. *Computational Intelligence*, *4*(4), 349-355.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., & Yan, H. (1999). Embodiment in conversational interfaces: Rea. *CHI'99: Proceedings of the SIGCHI conf. on Human factors in comp. syst.* (pp. 520-527). New York, NY, USA: ACM Press.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the sixteen personality factor questionnaire (16 PF)*. Champaign, Ill.: Institute for Personality and Ability Testing.

Costa, P. T., & McCrae, R. R. (1992). Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment*, *4*(1), 5-13.

Dastani, M. (2008). 2APL: a practical agent programming language. *Autonomous Agents and Multi-Agent Systems*, *16*(3), 214-248.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319-340.

Dias, J., & Paiva, A. (2005). Feeling and Reasoning: A Computational Model for Emotional Characters. In C. Bento, A. Cardoso, & G. Dias (Eds.), *Progress in Artificial Intelligence* (Vol. 3808, pp. 127-140). Berlin, Heidelberg: Springer Berlin Heidelberg.

Doce, T., Dias, J., Prada, R., & Paiva, A. (2010). Creating Individual Agents through Personality Traits. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent Virtual Agents* (Vol. 6356, pp. 257-264). Berlin, Heidelberg: Springer Berlin Heidelberg.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*, Series in affective science (pp. 572-595). New York, NY, USA: Oxford University Press.

Evertsz, R., Ritter, F. E., Busetta, P., & Pedrotti, M. (2008). Realistic Behaviour Variation in a BDI-based Cognitive Architecture. *Proc. of SimTecT'08*. Presented at the SimTecT '08, Melbourne, Australia.

Eysenck, H. J. (1976). The measurement of personality. *Proc. Royal Society of Medecine*, *40*(2), 75-80.

Eysenck, H. J., & Eysenck, S. B. G. (1968). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: EdITS.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, *6*(1), 21-29.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Frijda, N. H. (2006). *The Laws of Emotion*. Psychology Press.

Gillis, P. D., & Hursh, S. R. (1999). Using behavior moderators to influence CGFcommand entity effectiveness and performance. *Proceedings of the Eighth Conference on Computer Generated Forces and Behavioral Representation* (pp. 237-251). Orlando, FL, USA.

Goldberg, L. R. (1981). Language and individual differences: The search for universal in personality lexicons. (L. Wheeler, Ed.)*Review of personality and social psychology*, *2*(1), 141-165.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216-1229.

Goldberg, L. R. (1992). The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, *4*(1), 26-42.

Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th ed.). New York, NY, USA: Oxford University Press.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, *5*(4), 269-306.

Hayes-Roth, B., Brownston, L., & van Gent, R. (1995). Multiagent collaboration in directed improvisation. *Proceedings of the First International Conference on Multi-Agent Systems* (pp. 148-154). Presented at the ICMAS-95, San Francisco, CA, USA.

Howden, N., Rönnquist, R., Hodgson, A., & Lucas, A. (2001). JACK intelligent agents-summary of an agent infrastructure. *Proc. of the 5th International Conference on Autonomous Agents*.

Huber, M. J. (1999). JAM: a BDI-theoretic mobile agent architecture. *Proceedings of the third annual conference on Autonomous Agents*, AGENTS '99 (pp. 236–243). Seattle, Washington, USA.

Jiang, H., Vidal, J. M., & Huhns, M. N. (2007). EBDI: an architecture for emotional agents. *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, AAMAS '07 (pp. 11:1–11:3). Honolulu, Hawaii: ACM.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big Five Trait Taxonomy: History, Measurements, and Conceptual Issues. *Handbook of Personality: Theory and Research* (3rd ed.). The Guilford Press.

John, O. P., Robins, R. W., & Pervin, L. A. (Eds.). (2008). *Handbook of Personality: Theory and Research* (3rd ed.). The Guilford Press.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artif. Intell.*, *33*(1), 1-64.

Lim, M. Y., Dias, J., Aylett, R., & Paiva, A. (2008). Improving Adaptiveness in Autonomous Characters. In H. Prendinger, J. Lester, & M. Ishizuka (Eds.), *Intelligent Virtual Agents* (Vol. 5208, pp. 348-355). Berlin, Heidelberg: Springer Berlin Heidelberg.

Maes, P. (1994). Agents that reduce work and information overload. *Commun. ACM*, *37*(7), 30-40.

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

Malatesta, L., Caridakis, G., Raouzaiou, A., & Karpouzis, K. (2007). Agent personality traits in virtual environments based on appraisal theory predictions. *Proceedings of AISB'07: Artificial and Ambient Intelligence, Language, Speech and Gesture for Expressive Characters*. Newcastle, UK.

Maslow, A. H. (1968). *Toward a psychology of being* (2nd ed.). Princeton, NJ: Van Nostrand.

McAdams, D. P. (1992). The Five-Factor Model In Personality: A Critical Appraisal. *Journal of Personality*, *60*(2), 329-361.

McRorie, M., Sneddon, I., Sevin, E., Bevacqua, E., & Pelachaud, C. (2009). A Model of Personality and Emotional Traits. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Vilhjálmsson (Eds.), *Intelligent Virtual Agents* (Vol. 5773, pp. 27-33). Berlin, Heidelberg: Springer Berlin Heidelberg.

Neill, J. (2008). Personality Reading: Ch 15. *Center for Applied Psychology*. Retrieved from http://wilderdom.com/JamesNeill.htm

Norling, E., & Ritter, F. E. (2004). Towards Supporting Psychologically Plausible Variability in Agent-Based Human Modelling. *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems*. Presented at the AAMAS 2004.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology*,

Ortony, A. (2003). On making believable emotional agents believable. In R. Trappl, P. Petta, & S. Payr (Eds.), *Emotions in humans and artifacts* (pp. 189-211). Cambridge, MA: MIT Press.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotions* (Cambridge University Press.). Cambridge, UK.

Pelachaud, C. (2000). Some considerations about embodied agents. *Proc. of the Workshop on "Achieving Human-Like Behavior in Interactive Animated Agents"*. Presented at the The Fourth International Conference on Autonomous Agents, Barcelona, Spain.

Rao, A. S., & Georgeff, M. P. (1995). BDI Agents: From Theory to Practice. *Proc. of the 1st International Conference on Multi-Agent Systems* (pp. 312-319). Presented at the ICMAS-95, San Francisco, CA, USA.

Rizzo, P., Veloso, M., Miceli, M., & Cesta, A. (1997). Personality-Driven Social Behaviors in Believable Agents. *Proceedings of the AAAI Fall symposium on Socially Intelligent Agents* (pp. 109-114).

Rousseau, D., & Hayes-Roth, B. (1996). *Personality in synthetic agents* (Technical Report No. KSL 96-21). Stanford University, CA: Knowledge Systems Laboratory.

Russell, S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice Hall.

Sansonnet, J.-P., & Bouchet, F. (2010). Expression of behaviors in Assistant Agents as influences on rational execution of plans. In J. M. Allbeck, N. I. Badler, T. W. Bickmore, C. Pelachaud, & A. Safonova (Eds.), *Intelligent Virtual Agents, 10th International Conference*, Lecture Notes in Computer Science (Vol. 6356, pp. 413-419). Presented at the IVA 2010, Philadelphia, PA, USA: Springer.

Sardina, S. (2007). *A Course on BDI architectures*. University of Melbourne, Australia.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, *76*(4), 613-627.

Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006). Human Behavior Models for Agents in Simulators and Games: Part I: Enabling Science with PMFserv. *Presence: Teleoperators and Virtual Environments*, *15*(2), 139-162.

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, *43*(1).

Weiss, G. (2000). *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA: MIT Press.

Wooldridge, M. J. (2000). *Reasoning about rational agents*. MIT Press.

KEY TERMS & DEFINITIONS

Intelligent agent: an autonomous entity in interaction with its environment, from which it collects data through sensors, and upon which it acts through actuators, in order to reach a set of personal goals.
Rational agent: agents that rely exclusively on rational choice theory to decide its next set of actions, as opposed to an agent trying to mimic human non-rational behaviors (emotions, personality traits…).
Conversational agent: an intelligent agent in interaction with human users, usually in a multimodal way including at least natural language (typed or speech as input, displayed on screen or vocalized as output), and possibly gesture or emotions (if the agent has an embodiment).
Gloss: short phrase used in a thesaurus (or an electronic lexical database like WordNet) to define a word (or a synset).
Personality traits: a set of basic dispositions that influence the actions and behaviors of a person or agent, and that are stable over time (as opposed to moods or emotions).
Behavioral scheme: the smallest unit of concrete behavior that can be inferred from a personality trait.
Influence operator: symbolic rule-based heuristics handled by the behavioral engine (within the R&B framework), in order to alter the rational behavior of the agent, to make it non-exclusively rational.