Reference:
Sebastian Ahrndt, Frank Trollmann, Johannes Fähndrich, and Sahin Albayrak. Personality and agents: Formalising state and effects. In Matthias Klusch, Rainer Unland, Onn Shehory, Alexander Pokahr, and Sebastian Ahrndt, editors, 14th German Conference on Multiagent System Technologies (MATES 2016), number 9872 in Lecture Notes in Artificial Intelligence, pages 18–26. Springer International Publishing, 2016

Personality and Agents: Formalising State and Effects

Sebastian Ahrndt, Frank Trollmann, Johannes Fähndrich, and Sahin Albayrak

DAI-Laboratory of the Technische Universität Berlin
Faculty of Electrical Engineering and Computer Science
Ernst-Reuter-Platz 7, 10587 Berlin, Germany
sebastian.ahrndt@dai-labor.de (Corresponding author)

Abstract. Personality is one of the central elements determining the behaviour of humans. It influences other cognitive mechanisms such as emotions and moods and thus effects attention and actions. However, in the literature about cognitive agents, work that investigates the effects of personality is rare and somewhat disconnected. Bridging this gap represents one step towards conceptualising human behaviour in software agents, e.g. for resource-bounded agents in highly-dynamic environments or for virtual humans with realistic behaviour. The integration of personality in agents also requires its integration into reasoning processes used in agent-based systems. In this paper, we propose a formalisation that enables reasoning about the effects and state of personality. This formalisation is integrated into the '$\mathcal{L}$ogic $\mathcal{O}$f $\mathcal{R}$ational $\mathcal{A}$gents' ($\mathcal{LORA}$) and is the foundation for reasoning about the personality of other agents and the influence of personality on the action selection process.

Keywords: User/Machine Systems, Human factors, Software psychology, Cognitive models, Logic-based approaches and methods

1 Introduction

While reading about cognitive characteristics in agent-based system, one will soon recognise several approaches that bring emotions to artificial agents. Available work reaches from modelling and applying emotions [10, 11] to (completely axiomatised) logics of emotions [1, 7, 18]. The latter enable discussion and analysis of the effects of emotions on decision-making in a use-case independent and principle manner.

The integration of personality has not been studied in as much detail (*cf.* Section 2). However, following [14], personality is a significant factor for human behaviour and determines the outcome of essential behavioural processes, e.g., cognition and emotional reactions. Furthermore, it influences other affective phenomena such as moods and thus should be a central element in the reasoning and deliberation process of cognitive agents. The foundation of these processes is a formalism that can represent personality and can be integrated into existing

formalisms used in the reasoning processes of cognitive agents. Within this work, we propose a formalisation of personality that enables reasoning about the state and effects of personality.
We begin our discussion by reviewing existing approaches that integrate personality in agent-based systems in Section 2 and comparing them to approaches for formalising emotions. Afterwards, we discuss the objectives of this work in detail. In Section 4 we first describe $\mathcal{LORA}$ in an abstract manner to provide the necessary information to comprehend the remaining parts of the work and further present our integration of personality. Section 5 provides final remarks and insights into future work.

2 Related Work

Excursus – Personality Theories: Human factor psychology describes a human's personality by means of traits or types. What these approaches have in common is that traits or types are characteristic features of human beings, and that the human's behaviour and motives can be explained along behavioural patterns. Today, there are two well-established theories about human personality, namely: the *Five-Factor Model* of personality and the *Myers-Briggs Type Indicator*. We discussed the differences between both approaches and showed that the former is more suited for the integration of personality in agents in [2].

The representation of cognitive characteristics has a long tradition in agent-based systems and reaches from individual characteristics such as emotions, moods and personality to complex behavioural influences such as the cultural background. The areas of interest include next to virtual agents, virtual humans and personal assistants also (multi)agent-based simulations in different domains (e.g. traffic simulation, or crowd simulation). Another branch of research focuses on modelling and examining the effects of personalities on interactions between agents and their environments. In particular, the effects of personalities in cooperative settings. We have published a comprehensive analysis of available state of the art in prior work [3] that shows that the implemented effects of personality are often specific for the considered use-case and not applicable in general.
Most approaches define the effects/influences of personality in a rule-based or scripted manner. Unfortunately, personality traits are not inherently good or bad; their influence is context-dependent. This makes reasoning about personality influences a problem which is hard to solve by rule-based approaches. [17] presents a more advanced agent-based model of personality based on MBTI. In [2, 3], we showed how this can be achieved for the FFM, extending [8] to the complete set of personality traits.
Although these approaches consider personality in isolation, the overall objective is to build an agent-model that brings together all cognitive characteristics. This is discussed in some work, for example: [11] presents an architecture for artificial characters with personality, emotions, and moods based on the BDI model;

[10] proposes an extension of the BDI model introducing information about the personality, emotions and the physiology of a human; [5] introduces a concept for virtual characters that include specific personalities and emotional reactions. These approaches discuss architectural considerations from the software engineering perspective and provide first steps towards the integration of more than one cognitive characteristic.
A variety of approaches formalise emotions in agents. The PhD project of *Carole Adam* [1] provides an in-depth analysis of this topic and proposes a logic of emotions in agents. The formalisation is based on the OCC model and realised by expressing emotions based on the modalities beliefs, desires, and intentions. For instance, joy is defined as a feeling that happens when an agent is pleased about a desirable event.[1] Analogously, [7] presents an approach that formalises the intensity of emotions using the concept of graded modalities. The PhD project of *Bas R. Steunebrink* [18] describes a complete framework that formalises the emotional reactions from appraisal to coping.
Comparing the work on emotion and personality in agents reveals a gap with respect to theoretical and practical maturity. Our long-term goal is to bridge this gap. Our first step is a formalisation of the concept of personality. In contrast, to logic of emotions, this can not be done by using combinations of existing modalities, but via a new modality (*cf.* Section 4.1).

3 Objective: Reasoning about Personality

The objective of our work is to enable (1) reasoning about the influence of personality on the behaviour of an agent and (2) to derive the characteristics of personality of an agent from observations of its' behaviour. These two aspects requires a formalism that is able to represent (a) the effect of personality on the behaviour of an agent and (b) the state of personality of an agent. The representation of this formalism is the main contribution of this paper.
Reasoning about the influence of personality and deriving personality characteristics requires an underlying representation of personality, its effect and its state. Representing the effect of personality requires the ability to formulate that sth. is derived from the agents personality. Using commitment-strategies as example we could make the following informal statement:

1. Due to its personality agent i tends towards open-minded commitment.

This statement specifies that the cause for the open-minded commitment is the agents personality (as opposed to other factors such as incomplete knowledge). Such statements can be used to reason about consistent behaviour that is explained by the agents' personality.
Further necessary statements take into account the reasoning about personalities and the comparison of personalities of agents. Here statements of the following form are interesting:

[1] Using the syntax of $\mathcal{LORA}$ this could be formalised as $(Joy\ i\ \varphi) = (Bel\ i\ \varphi) \wedge (Des\ i\ \varphi)$ [1, p. 100 – 101].

2. Agent i is conscientious.

In addition to discrete classes of personality it can be usefull to talk about the extent of personality traits:

3. The agent i is more conscientious than agent k.
4. The agent i is very conscientious.

Finally, both kinds of statement can be combined to formulate dependencies between the personality of an agent and its behaviour:

5. If agent i is very conscientious then due to its personality agent i tends towards open-minded commitment.

This statement is an implication built from statements 4 and 1. It can be used to derive the agents' behaviour, when the personality state is known (1). In the example it would be possible to derive that the agent is likely to be open-minded when it is conscientious. The statement can also be used to derive information about personality traits from the agents behaviour (2). In the example, an agent that does not exhibit open-minded commitment is less likely to be conscientious.

4 Formalisation of Personality

This section presents our integration of personality into $\mathcal{LORA}$ [20]. We start with a short introduction into the logic. $\mathcal{LORA}$ is a logic developed to enable reasoning about the behaviour of agents using the modality operators *Belief*, *Desire*, and *Intention*. The vocabulary is based on the sorts Ag – agents, Ac – actions, Gr – groups of agents, and U – other individuals. For these sorts constants and variables can be defined and used in first-order predicate logic formulas together with additional domain-specific formulas. A model in the logic contains a domain description $D = \langle D_{Ag}, D_{Ac}, D_{Gr}, D_U \rangle$, specifying the available entities for each sort. A temporal dimension is added by a set of time points T and a branching, temporal relation $\mathcal{R} \subseteq T \times T$. As the logic follows a possible world semantic, a set of worlds W is defined. Worlds are related to temporal structures. The operators $\mathcal{B}$, $\mathcal{D}$ and $\mathcal{I} : D_{Ag} \to \wp(W \times T \times W)$ represent the modalities. They map an agent to a set of triples, each of which assigns a possible world and a time point to another world. To reason about belief accessible worlds the shortcut function $\mathcal{B}_t^w(i) = \{w' | \langle w, t, w' \rangle \in \mathcal{B}(i)\}$ is defined. It denotes that agent i beliefs that w' is a possible world state at time point t. Shortcut functions for $\mathcal{D}$ and $\mathcal{I}$ are defined analogously. The elements in the sets $\mathcal{B}/\mathcal{D}/\mathcal{I}_t^w(i)$ represent i in world w at time point t.
Based on the possible world definition of these three modalities reasoning about the fulfilment of formulas is implemented. For $\mathcal{B}$ this can be done via $(Bel\ i\ \varphi)$ stating that an agent i believes φ. These state formulas are evaluated with a specific world w and a specific time point t. The formula Bel is defined as: $\langle w, t \rangle \models (Bel\ i\ \varphi)$ iff $\forall w' \in B_t^w([\![i]\!]), \langle w', t \rangle \models \varphi$, where $[\![i]\!]$ is the evaluation of term i under a variable assignment. Intuitively, this statement states that an agent i believes a statement φ if this statement holds in all worlds that are accessible via its beliefs.

4.1 Representing the Effects of Personality

In this section, we discuss how to extend $\mathcal{LORA}$ to enable reasoning about the influence of personality on the behaviour of an agent. This is done by introducing personality as new modality. This design decision was made based on the observation that the influence of emotions is frequently represented by combinations of the existing modal operators (Believes, Desires and Intentions) (cf. [1, 7, 18]). However, personality is different from emotion as it "...is the coherent patterning of affect, behavior, cognition, and desires (goals) over time and space" [16]. In contrast, the effects of emotions are bounded to a particular time and object [13]. In fact, emotions always occur relative to something (object, event, action) [12]. "A helpful analogy is to consider that personality is to emotion as climate is to weather" [16]. Thus, psychologist consider personality to be (to some extent) a time and space independent cognitive mechanism; that influences each stage of the decision-making process of humans [16]. To substantiate this statement the interested reader is referred to work that shows that we as humans have a relatively stable personality over our lifespan as adults (*cf.* [6, 9, 19]).
The conclusion we draw from these findings is that personality is per se independent of the *Beliefs*, *Desires*, and *Intentions* at specific times. Therefore, our approach is to represent personality as dedicated modality. We have presented an extension of the syntax and semantics of $\mathcal{LORA}$ introducing a personality modality in prior work [4]. The effects of personality of each agent are given by the modality operator $\mathcal{P} : D_{Ag} \rightarrow \wp(W \times T \times W)$. The operator $\mathcal{P}$ is named personality-accessibility relation. It defines all worlds that are in line with the personality of an agent $i \in D_{Ag}$ given a specific situation $\langle w, t \rangle$, where $w \in W$ and $t \in T$. Analogous to the other modalities the shortcut function $\mathcal{P}_t^w(i) = \{w' | \langle w, t, w' \rangle \in \mathcal{P}(i)\}$ can be used to reason about personality in a specific world and time. Reasoning about the fulfilment of formulas is enabled by a state formula (Per $\langle$*ag-term*$\rangle$ $\langle$*state-fmla*$\rangle$). The semantics of this state formula is defined based on an agent i and a state formula φ. We assume that the statements are usually evaluated in the context of a fixed model and variable assignment and omit them for the sake of brevity. Consequently, the semantic of Per is defined as $\langle w, t \rangle \models (Per\ i\ \varphi)$ iff $\forall w' \in \mathcal{P}_t^w([\![i]\!]), \langle w', t \rangle \models \varphi$. The formula (Per i φ) verbalises the fact that agent i tends to φ. Here *tends to* refers to the influence of personality and not to other preferences, e.g., in terms of emotions and moods. Indeed, *tends to* is a placeholder for a personality-descriptive verb that must be used in a specific situation [15]. In a general manner it can be interpreted as φ being aligned with the personality of agent i, as this formula holds in all worlds which are accessible with the personality of i.
This operator can now be used to describe the first statement from our objectives. We can say that due to its personality agent i tends towards open-minded commitment via the state-formula:

$$(\text{Per}\ i\ hasOpenMindedCommitment(i))$$

Here, $hasOpenMindedCommitment(i)$ is a predicate describing that an agent i has open-minded commitment.

4.2 Representing the State of Personality

Reasoning about the state of personality of an agent requires having a notion to represent this state. As described in Section 2, we consider the FFM to be most suited for the integration of personality in agents. In FFM each personality trait is represented by a continuous scale.[2] Hence, one personality consists of a real number value for each trait. These values are interpreted with respect to a maximum and minimum (e.g., 1 and -1), where the maximum means that the factor is fully developed, the minimum means that the factor is not developed, and the average means that the factor is balanced. For example, a value of 1 for extraversion denotes that the person is considered extroverted while a value of -1 means the person is introverted and a value of 0 means that neither a strong tendency towards introversion nor extravorsion can be observed.

To include this model into $\mathcal{LORA}$ we first need to enable handling real numbers to express and compare the extent of personality traits. For this purpose the comparison functions $=$, $<$ and $>$ can be used. These are integrated as additional state formulas comparing two real number expressions: $\mathbb{R} = \mathbb{R}$; $\mathbb{R} < \mathbb{R}$; $\mathbb{R} > \mathbb{R}$. For further use cases other real-valued expressions (e.g., addition or multiplication) may be relevant. These can be integrated analogous to the statements above.

For the formalisation, we assume that the personality only depends on the agent itself and is stable over time. Thus, the personality does not depend on the world or the time point but solely on the agent. To represent the state of personality we define one function per personality factor that maps the agent to the value representing the extent of the respective personality trait: $O, C, E, A, N : Ag \to \mathbb{R}$. The numbers derived from personality traits usually need to be interpreted in some way. For instance, on the scale presented above (-1 to 1) it could make sense to exclude personalities between -0.3 and 0.3 as they may be considered to be roughly balanced. For the trait extraversion the two extremes of the scale can be interpreted as introversion and extroversion. Here we could consider agent i to be introverted if $E(i) < -0.3$ and extroverted if $E(i) > 0.3$ and neither of those if $-0.3 < E(i) < 0.3$. This enables discrete reasoning about personality categories as in MBTI but allows for the definition of more nuanced subclasses. Constants and variables representing real numbers are required to express such statements. Those can be integrated into $\mathcal{LORA}$ analogously to the variables and constants of other sorts, e.g., variables denoting agents. For readability we denote constants by their actual values, e.g., 0.3 is a constant of value 0.3 whose name is "0.3".

These statements now enable expressions that refer to personality traits of agents and interpret them, either in the context of personality traits of other agents or in the context of variables or constants. They are sufficient to express statements 2 to 4. Statement 2 denotes that an agent i is conscientious. It could be expressed as $C(i) > 0.3$. Comparing two agents conscientiousness, e.g., to state that an agent i is more conscientious than an agent k (statement 3) can be expressed as

[2] We will use the following abbreviations next: openness to experience (O), conscientiousness (C), extraversion (E), agreeableness (A) and neuroticism (N).

follows: $C(i) > C(k)$. Intervals in the continuous personality scale can be used to express more fine-grained personality trait distinctions. Here, we could consider an agent to be very conscientious when it has a higher trait than 0.8. Using this (arbitrary) line we can formulate statement 4 as $C(i) > 0.8$.

5 Discussion and Future Work

The goal of our formalisation is to enable reasoning about the interdependencies of personality and behaviour of a natural agent. Section 4 describes how $\mathcal{LORA}$ can be extended to represent the effects and state of personality. Both extensions can be combined to express how specific personality types (*i.e.*, the state of personality) influence the agent. An example is given in the fifth statement in Section 3, which expresses that a very conscientious agent tends towards open-minded commitment. This can be expressed as follows:

$$C(i) > 0.8 \rightarrow (\text{Per } i \; hasOpenMindedCommitment(i)).$$

Such statements represent the relation between state and effects of personality and can be used for reasoning, e.g., along the implication operator. The extension of $\mathcal{LORA}$ can also be used for further discussions about the relation of the formal representation of personality to other parts of the logic.

Integrating personality as own modality provides the fundamentals for a comprehensive analysis of the properties that are useful to characterise an agent with personality. From a purely logical aspect there is no reason to do that, as we have build a formal system enabling us to use all possible combinations of the formulas and operators available. However, that would mean to ignore the semantic of the properties, *i.e.*, to not discuss how the properties influence the behaviour of an agent and which influences are meaningful/reasonable for analysing personality driven behaviour of agents. A full discussion of these relations will be done in future work.

Another relation that could be beneficial to observe is the relation between formalisations of emotions and our formalisation of personality. Several authors represent emotions via formulas over the believe, desire and intention modalities. Personality also influences the way in which we react to situations emotionally, *i.e.*, the occurrence, intensity, and duration of emotions. The representation of personality presented in this paper provides a foundation for considering such relations between personality and emotions.

Although our extension enables the integration of personality into the reasoning process in general, it does not enable to derive relation between personality and behaviour directly. Doing so in a general way requires formalising findings from psychology in the form of statements that can be used for reasoning among multiple approaches. Our formalisation provides a vocabulary to express those statements.

References

1. Adam, C.: Emotions: from psychological theories to logic formalization and Implementation in a BDI agent. L' Institute National Polytechnique de Toulouse (2007)
2. Ahrndt, S., Aria, A., Fähndrich, J., Albayrak, S.: Ants in the OCEAN: Modulating agents with personality for planning with humans. In: Bulling, N. (ed.) Multi-Agent Systems, Revised Selected Papers, pp. 3–18. No. 8953 in LNAI, Springer (2015)
3. Ahrndt, S., Fähndrich, J., Albayrak, S.: Modelling of personality in agents: From psychology to implementation. In: Bordini, Elkind, Weiss, Yolum (eds.) Proc. of HAIDM 2015, co-located with AAMAS 2015, pp. 1–16. IFAAMAS (2015)
4. Ahrndt, S., Fähndrich, J., Lützenberger, M., Albayrak, S.: Modelling of personality in agents: From psychology to logical formalisation and implementation. In: Bordini, Elkind, Weiss, Yolum (eds.) Proc. of AAMAS 2015, pp. 1691–1692. IFAAMS (2015)
5. Bevacqua, E., de Sevin, E., Pelachaud, C., McRorie, M., Sneddon, I.: Building credible agents: Behaviour influenced by personality and emotional traits. In: Lèvy, Bouchard, Yamanaka, Aoussat (eds.) Proc. of KEER 2010. pp. 1–10 (2010)
6. Caspi, A., Roberts, B., Shiner, R.: Personality development: Stability and change (56), 453–484 (2005)
7. Dastani, M., Lorini, E.: A logic of emotions: from appraisal to coping. In: Conitzer, Winikoff, Padgham, van der Hock (eds.) Proc. of AAMAS 2012. IFAAMAS (2012)
8. Durupinar, F., Allbeck, J., Pelechano, N., Badler, N.: Creating crowd variation with the OCEAN personality model. In: Padgham, Parkes, Müller, Parsons (eds.) Proc. of AAMAS 2008, pp. 1217–1220. IFAAMAS (2008)
9. Hampson, S.E., Goldberg, L.R.: A first large-cohort study of personality-trait stability over the 40 years between elementary school and midlife (91), 763–779 (2006)
10. Jones, H., Saunier, J., Lourdeaux, D.: Personality, emotions and physiology in a bdi agent architecture: The pep ->> bdi model. In: Proc. of IAT 2009, pp. 263–266. IEEE (2009)
11. Neto, A.F.B., da Silva, F.S.C.: On the construction of synthetic characters with personality and emotion. In: da Rocha Costa, A.C., Vicari, R., Tonidandel, F. (eds.) Advances in Artificial Intelligence SBIA 2010, LNCS, vol. 6404, pp. 102–111. Springer (2010)
12. Oatley, K., Jenkins, J.M.: Understanding Emotions. Blackwell Publishing (1996)
13. Ortony, A., Norman, D., Revelle, W.: Affect and proto-affect in effective functioning. In: Fellous, J.M., Arbib, M.A. (eds.) Who needs Emotions?, pp. 173–202. Series in Affective Science, Oxford University Press (April 2005)
14. Ozer, D.J., Benet-Martínez, V.: Personality and the prediction of consequential outcomes. Annu. Rev. Psychol. 57, 401–421 (2006)
15. de Raad, B., Mulder, E., Kloosterman, K., Hofstee, W.K.: Personality-descriptive verbs. European Journal of Personality 2, 81–96 (1988)
16. Revelle, W., Scherer, K.R.: Personality and emotion. In: Sander, D., Scherer, K. (eds.) The Oxford Companion to Emotion and the Affective Sciences, p. 512. Series in Affective Science, Oxford University Press (2010)
17. Salvit, J., Sklar, E.: Modulating agent behavior using human personality type. In: Proc. of HAIDM at AAMAS 2012. pp. 145–160 (2012)
18. Steunebrink, B.: The Logical Structure of Emotions. Utrecht University (2010)
19. Wilks, L.: The stability of personality over time as a function of personality trait dominance. Griffith University Undergraduated St Psy J 1, 1–9 (2009)
20. Wooldridge, M.: Reasoning about Rational Agents. Intelligent Robotics and Autonomous Agents, The MIT Press (July 2000)