Context-aware Agents based on Psychological Archetypes for Teamwork

Extended Abstract

Anupama Arukgoda
University of New South Wales
Canberra, Australia
a.arukgoda@adfa.edu.au

Erandi Lakshika
University of New South Wales
Canberra, Australia
e.henekankanamge@adfa.edu.au

Michael Barlow
University of New South Wales
Canberra, Australia
m.barlow@adfa.edu.au

Kasun Gunawardana
University of Colombo School of Computing
Colombo, Sri Lanka
kgg@ucsc.cmb.ac.lk

ABSTRACT

As artificial agents become ubiquitous with the advancements of Artificial Intelligence (AI), creating effective, viable artificial teammates has become increasingly important. Multiple research studies have attempted to personify agents endowing them with different archetypes observed in human psychology theories with the aim of creating realistic, predictable and believable agents. However, when these agents are exposed to other agents (both artificial and human), the archetypal qualities should be amenable to create socially believable and socially intelligent agents. This paper presents a generic framework to model personified archetypes of agents. The framework provides a flexible platform that can accommodate the behavioral changes of an agent influenced by many contextual factors. The proposed framework will drive better designing of effective believable synthetic agents/characters and more user-friendly virtual assistants customized to a human's personality.

KEYWORDS

Context-aware agents; Archetypes; Goal-driven agents

ACM Reference Format:

Anupama Arukgoda, Erandi Lakshika, Michael Barlow, and Kasun Gunawardana. 2023. Context-aware Agents based on Psychological Archetypes for Teamwork: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

To be qualified as good teammates, artificial agents should be embedded with qualities beyond objective performance [1, 18]. Factors such as reliability, trust, predictability, which are often challenging to engineer, are required for humans to perceive agents as better assistants, teammates and companions [10, 19]. Many research studies have explored the personification of agents with human attributes to create more realistic, approachable agents [3, 9]. While endowing agents with characteristic behavior is useful, intelligent agents should also be capable of revising their behavior in response to their context [6]. Hence, we propose a general framework to model archetypes of agents whose innate behavior is amenable to the context with the aim of creating realistic, socially adequate and personified agents.

We define an archetype as an agent with an innate set of objectives according to which they define their preferences over goals. The objectives could be manifested as motives, values, dispositions or even as utility functions the agent tries to optimize. While there already exists different frameworks to model motivation [5, 20], personality traits [4], desires and values, our framework abstracts them as archetypes so that our framework can make future research consistent.

Psychological studies discuss phenomena such as conformity, divergence, and vicarious living, that emerge when individuals are exposed to social settings [2, 8]. However, a detailed analysis of the implications of social context on archetypes and rich team dynamics it could lead to, is yet to be explored [13]. The proposed framework accommodates the malleability of agent archetypal behavior and enables future researchers to experiment psychology and sociology theories on one platform.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

2 PROPOSED FRAMEWORK

The proposed framework provides a versatile platform in comparison to previous work where certain archetypal behaviours have been modelled using mathematical equations [15, 16], constraining their adaptability to context. Our framework abstracts any goal selection behaviour differences proposed in previous literature, such as motives [5, 20], personalities [4], emotional differences [5, 12] etc., as archetypes. This makes our framework a generalized, common platform to conduct such research, with the added benefit of context adaptability.

As shown in Figure 1, the framework comprises of 4 layers - Cognitive, Archetype, Drive Elicitor and Goal Scoring layer. An agent acquires information via its sensors, which helps the agent formulate and revise its contextual information and generate a description for each potential goal. The perceived social context will inadvertently revise the agent's archetypal behaviors/drives. Simultaneously, the archetypal drives evoked by the goal are identified within the Drive Elicitor layer. Finally, the agent's desirability towards the goal is calculated based on the difference between the archetypal drives that have been revised and the drives energized by

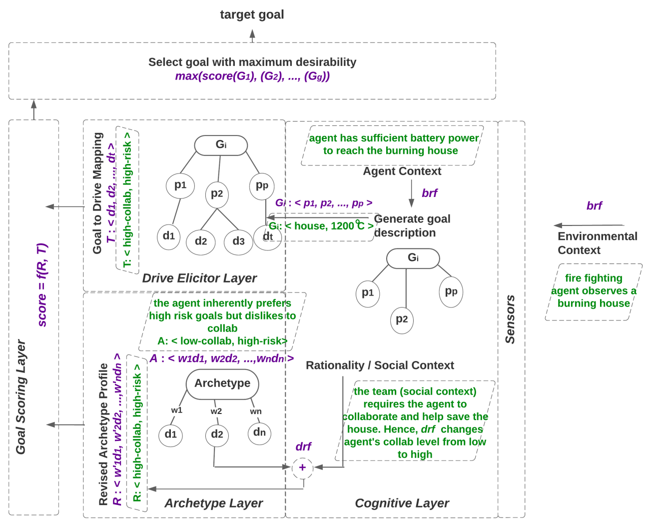

Figure 1: Proposed framework for context-aware archetypes

the current goal. Potential techniques to implement each layer and how an archetype would select a goal reasoned by its inherent characteristics are discussed. In Figure 1, we have also depicted within dotted parallelograms a running example of how a fire-fighting agent who inherently prefers taking high risks but dislikes collaboration (collab) assigns a desirability score to a burning house which needs to be saved collaboratively.

2.1 Cognitive Layer

The Cognitive Layer processes the information about the agent's environment (Environmental Context - e.g. burning house) and about its own self (Agent Context - e.g. agent has sufficient battery). At this layer an agent identifies potential goals based on the information state such as beliefs about the external environment obtained through its one or more sensors. This layer stores and processes the agent's own constraints and capabilities, beliefs, and view of the external environment using belief-revision-functions (***brf***) that could affect its perception of a goal. Beliefs of a rational agent could be revised based on its perception, communication and contemplation as proposed by Jiang et al. [12]. Hence, this layer will output the agent's definition of a goal based on its perception. A potential goal G_i is defined by its properties. Hence, it can be represented by a p-tuple where $p_1 - p_p$ are the outputs of the agent's sensors; $G_i : < p_1, p_2, ..., p_p >$ (e.g. $< house, 1200°C >$). Agent's beliefs get revised in this layer based on the frequency at which the agent re-evaluates the environment and the rate at which the environment changes. Furthermore, the agent derives the social context that it is deployed in within this layer. Agent's team awareness, social identity and agent organization, and trust towards the agents within its vicinity [14] are factors that affect the social context.

2.2 Archetype Layer

An archetype can be formally defined as a combination of its drives/preferences and the respective intensity levels of these drives. A multi-dimensional archetype can be represented as a n-tuple. $A : < w_1d_1, w_2d_2, ..., w_nd_n >$, where d_1-d_n represent n different drives and w_1-w_n represent their weights/intensity levels. The inherent archetypal drives could be influenced by the social context. For instance, agents could display behavior antithetical to their characteristic behavior due to peer pressure or even by the mere presence of another agent [7, 8]. Furthermore, this layer can also facilitate temporal changes to archetypes [13] and the plasticity of personality traits [11]. This adaptation/revision is represented using a drive revision function/mechanism ***drf***. Introducing new drives, exciting or inhibiting the intensity of the current drives and suppressing the effect of certain drives are the functionalities envisioned for ***drf***. Once ***drf*** manipulates the archetypal drive, the revised archetype profile takes the form $R : < w'_1d_1, w'_2d_2, ..., w'_nd_n >$.

Ideally, rationality defines objective, unbiased behavior. However, what is rational could also depend on the social context as what is socially appropriate may not be an unbiased action [17]. Hence, ***drf*** could be triggered by both rationality and social context. This layer essentially facilitates the creation of socially intelligent agents.

2.3 Drive Elicitor Layer

This layer is a domain specific and agent-specific layer. It is the middleware that maps the sensor outputs to the drives elicited by the goal properties based on agent's perception. The mapping could be predefined or implemented using techniques such as fuzzy logic. Output of this layer would take the form $T : < d_1, d_2, ..., d_t >$ assuming the sensor outputs are mapped to the drives d_1-d_t.

2.4 Goal Scoring Layer

This is the layer that finally calculates the agent's desirability towards a goal. This layer takes the goal-to-motive mapping (T) generated by the Drive Elicitor layer and revised archetype profile (R) from the Archetype layer. One mode of implementing this layer is to first map both inputs to utility values represented by real numbers (similar to a valance function) and then use a distance measure to calculate how much the goal aligns with the agent's archetypal preferences. Else, as with the example shown in Figure 1, generate a set of tags that define agent's archetype and the elicited drives by the goal and then calculate a document similarity measure such as overlap coefficient between R and T definitions adjusted by the weights of the archetypal drives.

This process is conducted for each potential goal, and then the goal with the highest desirability is selected as the target.

3 CONCLUDING REMARKS

This paper presents, a novel, generic framework to model context-aware agent archetypes. Personifying agents with archetypal attributes with our framework will enable the creation of believable, socially acceptable, and predictable synthetic agents. Such agents will enhance the experience of human interaction. Modelling socially intelligent archetypes and deploying them in collaborative tasks will indicate which archetype compositions work well together. Hence, endowing agents with archetypes will broach the subject of effective artificial, hybrid and even organic team designing, such that strengths and weaknesses of different archetypes synergize and complement well.

REFERENCES

[1] Gagan Bansal, Besmira Nushi, Ece Kamar, Eric Horvitz, and Daniel S Weld. 2021. Is the most accurate ai the best teammate? optimizing ai for teamwork. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11405–11414.

[2] Steven J Bosworth, Tania Singer, and Dennis J Snower. 2016. Cooperation, motivation and social balance. *Journal of Economic Behavior & Organization* 126 (2016), 72–94.

[3] Mercedes Bunz and Marco Braghieri. 2022. The AI doctor will see you now: assessing the framing of AI in news coverage. *AI & society* 37, 1 (2022), 9–22.

[4] André MC Campos, Frank Dignum, Virginia Dignum, Alberto Signoretti, Anne Magály, and Sérgio V Fialho. 2009. A process-oriented approach to model agent personality.. In *AAMAS (2)*. 1141–1142.

[5] José Cascalho, Luis Antunes, Milton Corrêa, and Helder Coelho. 2005. Toward a motivated BDI agent using attributes embedded in mental states. In *Conference of the Spanish Association for Artificial Intelligence*. Springer, 459–469.

[6] Célia da Costa Pereira, Andrea Tettamanzi, and Leila Amgoud. 2006. Goal revision for a rational agent. In *ECAI*. 747–748.

[7] Ayelet Fishbach and Melissa J Ferguson. 2007. The goal construct in social psychology. (2007).

[8] Ayelet Fishbach and Yanping Tu. 2016. Pursuing goals with others. *Social and Personality Psychology Compass* 10, 5 (2016), 298–312.

[9] Christopher Flathmann, Beau Schelble, Brock Tubre, Nathan McNeese, and Paige Rodeghero. 2020. Invoking Principles of Groupware to Develop and Evaluate Present and Future Human-Agent Teams. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 15–24.

[10] Adam J Hepworth, Daniel P Baxter, Aya Hussein, Kate J Yaxley, Essam Debie, and Hussein A Abbass. 2020. Human-swarm-teaming transparency and trust architecture. *IEEE/CAA Journal of Automatica Sinica* 8, 7 (2020), 1281–1295.

[11] Edmund R Hunt. 2020. Plastic 'personalities' for effective field swarms. In *Proceedings of The 3rd UK-RAS Conference*.

[12] Hong Jiang, Jose M Vidal, and Michael N Huhns. 2007. EBDI: an architecture for emotional agents. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. 1–3.

[13] Md Mohiuddin Khan, Kathryn Kasmarik, and Michael Barlow. 2018. Toward computational motivation for multi-agent systems and swarms. *Frontiers in Robotics and AI* 5 (2018), 134.

[14] Michael Macy and Andreas Flache. 2009. *Social dynamics from the bottom up: Agent-based models of social interaction*. Oxford University Press.

[15] Kathryn E. Merrick. 2016. *Computational Models of Motivation for Game-Playing Agents*. Springer. https://doi.org/10.1007/978-3-319-33459-2

[16] Kathryn E Merrick and Kamran Shafi. 2011. Achievement, affiliation, and power: Motive profiles for artificial agents. *Adaptive Behavior* 19, 1 (2011), 40–62.

[17] Diogo Rato and Rui Prada. 2021. Towards social identity in socio-cognitive agents. *Sustainability* 13, 20 (2021), 11390.

[18] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Thomas O'Neill, Richard Pak, and Moses Namara. 2022. Investigating the Effects of Perceived Teammate Artificiality on Human Performance and Cognition. *International Journal of Human–Computer Interaction* (2022), 1–16.

[19] Ho Chit Siu, Jaime Peña, Edenna Chen, Yutai Zhou, Victor Lopez, Kyle Palko, Kimberlee Chang, and Ross Allen. 2021. Evaluation of human-AI teams for learned and rule-based agents in Hanabi. *Advances in Neural Information Processing Systems* 34 (2021), 16183–16195.

[20] J Velsquez. 1997. Modeling emotions and other motivations in synthetic agents. *Aaai/iaai* (1997), 10–15.