The Source of Desire: Personal Identity as a Drive for Agent Behavior

Ursula Addison[1,*]

[1]CUNY Graduate Center, New York, NY, USA

Abstract

The central component in the design of our artificial agent behavior generation system, *la VIDA*, is its drive mechanism. This drive, which we have named an *identity profile*, is a collection of roles, values, beliefs, and attitudes analogous to and inspired by the human identity. This identity profile acts as an intrinsic source of motivation to drive long-term, autonomous agent behavior. In this paper we explore the representation and function of an early design for the identity profile. We also discuss how the identity profile interacts with a personality inventory and commonsense knowledge graph to assign traits and associated strengths to the agent. These values can then be mapped to three behavior intensities to augment agent actions.

Keywords

agential identity, artificial agent, artificial intelligence, cognition, intrinsic motivation, personality traits, behavior

1. Introduction

For an agent to achieve any degree of autonomy there must be something compelling their actions. One type of compulsion is motivation, i.e. a force that energizes, activates, and directs behavior. Motivation is as varied as the agents it compels. Deci and Ryan have identified two broad classes of motivation in humans, intrinsic and extrinsic [1]. Intrinsic motivation is activated by internal factors while extrinsic motivation arises from external factors.

la VIDA is a goal-driven autonomy (GDA) system, where a GDA system generates goals that become the focal point of system processes and agent actions. GDA systems typically have one or more drives which provide a sense of purpose to guide the agent. At the time of writing this paper we are unaware of any GDA system that uses intrinsic motivation as the sole drive type. Many GDA systems are given extrinsic drives, e.g. internal evaluation mechanisms or system designer specified achievement targets; Dora [2] and George [3] are systems of this variety. While we don't claim that our proposed method will be an improvement on existing approaches, we do claim that it could provide a unique perspective on autonomy. We hope that our work will inspire new questions and aid in finding their answers.

Behavior can be stimulated by both emotions and motivations, the key difference being that motivation directs and energizes behavior while emotion generally only energizes. Motivation is

AIC 2022, 8th International Workshop on Artificial Intelligence and Cognition, June 15–17, 2022, Örebro University, Örebro, Sweden

*Corresponding author.

uaddison@gradcenter.cuny.edu (U. Addison)

defined as *an urge that results in wants or needs that direct behavior toward a goal* [4]. Emotion is defined as *a conscious and subjective experience characterized by mental states, biological reactions, and expressions* [4]. Beyond the important fact that our primary interests lie with intrinsic motivation in GDA systems, emotions are less suited for our purposes due to their biological nature. This is not to say that our agents won't express emotions, but simply that emotions are not directly encoded in the identity profile. Several existing models incorporate intrinsic motivation using Reiss' 16 basic desires [5], the Five-Factor Model of Personality [6], or concepts from personality psychology [7]. Ramos et al. created a personality model using the 16 basic desires; personality traits are derived from each of the desires [8]. Cebolledo and Troyer apply Reiss' motivations in agents which seek to satisfy their own needs [9]. Sun described three types of drives; primary, high-level, and secondary. These drives are used to dynamically generate goals [10]. The contributions of this paper are as follows:

1. A novel conceptual and structural representation of an agent's roles, values, beliefs, and attitudes that is inspired by the human identity [7].
2. A novel GDA drive based on intrinsic motivation derived from an identity profile.
3. The design of a mapping between the personality traits found in the Mini-Marker Personality Inventory (PI) [11] and their empirically supported expected behavior. Each trait maps to three intensities of behavior, weak, medium, and strong.

This paper is organized as follows, section 2 provides an introduction to the foundational concepts of the identity profile and a short description of our system *la VIDA*, section 3 describes the identity profile design, and section 4 discusses some conclusions and future work.

2. Identity

When one refers to an identity they are essentially speaking about an agent's self-definition. The machinery provided by philosophy and psychology isn't always sufficient to describe the identity, but in general we consider it to encompass an autonomous agent's values, roles, beliefs, and personality. The agent uses these segments to describe themselves in past, present, and futures perspectives. They also use it to internally model their behavior and justify motivations [7].

Early work on identity asserted that it was fixed, however it is accepted today, that while a mature identity is very stable and results from both genetic and environmental factors, it is dynamic and changes throughout the stages of life. During an individual's life certain types of events such as crisis and commitment are key to shaping the identity [12]. Identity can be explored through the lens of a life account, and often begins to coalesce in adolescents who have a chance to experiment with different beliefs, roles, and values to find a combination that resonates. Key facets of a person's life account include: the body, individual characteristics, human relations, social reality, temporal environment, culture, and spiritual viewpoint. Just as an individual's internal self-image forms an integral part of their identity, so do the images others hold of them. These external images are internalized and integrated into the self-image.

There are a variety of ideas on how personhood is developed, but we focus on work by Williams James [13] that attributes at least part of identity formation to personal effort. Sustained

effort crystallizes a collection of ideas and experiences into a coherent reference. This reference "provides their lives with unity, purpose, and meaning" [7]. James maintained that the single identity is composed of three sub-identities; material self, social self, and spiritual self. The material self relates to aspects of the physical world including the agent's body. The social self relates to other agents and their images of the agent, as well as social norms, and internal values and roles. The spiritual self encompasses the agent's judgement, metaphysical beliefs, and spiritual beliefs.

An important alternative to James' work is the theory of identity formation based in personal narratives [7][14][15]. Through writing their life story, a person defines themselves within the narration. A life story may not necessarily be an accurate or complete portrayal of a life lived, but rather a story constructed by the author that is a meaningful compilation of experiences, thoughts, memories, and feelings used to paint a self-image.

2.1. Five-Factor Personality Model

A personality is the environmental and genetic impulse to interpret reality and respond to it in a certain manner. The Five-Factor Model (FFM) can capture a personality by weighting each of its five factors. The factors are *conscientiousness*, *agreeableness*, *neuroticism*, *openness*, and *extraversion* [6]. *Conscientiousness* refers to an agent's degree of precision and carefulness. *Agreeableness* impacts how likely an agent is to have a pleasant manner and conform under different circumstances. *Neuroticism* relates to mood stability and levels of anxiety. *Openness* determines the curiosity and adaptability of an agent. And finally, *extraversion* and its polar opposite introversion respectively correlate with an agent being more or less communicative and transparent. Each factor can hold a value within the interval $[0, 1]$, with a value of 1 linked to the most extreme version of a factor. In humans factor values are dynamic and are influenced by an individual's genes and environment.

2.2. Self-Schemata

A mental schema is a cognitive structure to manage information in the mind. When a schema exists for a particular dimension i.e. a target concept, new related information is grouped with existing information or applied to forming a new schema. The agent infers a conclusion about the dimension, and a generalization with respect to it is formed [16]. For example, money is needed to procure most goods and services. Clearly situations will arise when this claim isn't true, but it is a fairly safe and consistent assumption. The advantages of schema include understanding new knowledge by linking it to existing knowledge and simplifying vast amounts of information by focusing on a subset of it. Schemata allow agents to do more efficient processing of environments and events, even though there is no guarantee the output is correct. We are specifically interested in a sub-class of schema called self-schema, in which all information stored for a dimension is related to the agent [16]. For example, if the dimension is the role of working as a barista, self-schema information will be a combination of expectations about being a barista and the agent's beliefs, memories, and experiences in the role.

Schema can also be used to form behavioral scripts for commonplace situations. For instance, many people have expectations of how a cafe event would unfold, where the customer is

purchasing food or drinks. To minimize the use of cognitive resources in commonly experienced situations, an agent will often activate a behavioral script. Drawing information from the schema, the script models the expected scenario including all necessary objects, roles, and ordered events [17]. Schemata in combination with behavioral scripts have the potential to significantly reduce the processing needed to assess and act in various situations. We use concepts from self-schema, behavioral scripts, and the FFM personality model to create a cognitive structure to define the agent's identity.

2.3. Motivation

Motivation is the antecedent of a desire to be in a certain state. If an agent wants food, it may be motivated by thirst or hunger. In this situation, a possible goal state is to be full i.e. for the feeling of hunger to subside. Not all motivations are problems to solve or feelings that the agent wishes to recede, at times one wishes to increase and stimulate certain motivations. Regardless of the agent's attitudes toward any particular motivation, it can fuel action.

Motivation is a force that energizes, activates, and directs behavior toward a goal state [4]. Many motivations originate in the identity of an agent as a result of biological and social factors. The strength of each motivation is dynamic and an agent's set of motivations interact to drive the agent to act and exhibit behavior. Similar to other identity elements, motivations and their effects are subjective and there is no hard rule for the result of any particular configuration. By configuration we refer to the combination of motivations at play and their respective strengths. For our work we consider identity motivations in a highly simplified manner. This entails selecting a set of traits, where each is assigned a value in $[0, 1]$.

At this time, we are working with the personality traits found in the Mini-Marker PI [11]. We will then match each trait to the most commonly linked behaviors found in empirical research studies [18][19][20]. Taking guidance from the FFM, we assume there is expected behavior for each of the five factors depending on their value. An agent with a high *conscientiousness* value is expected to have precise and careful behavior. A low value in this factor is expected to result in careless and inattentive actions. In this way we capture motivations by mapping trait values to three intensities of anticipated types of behavior. We apply and represent motivation by starting with the identity component, link it to expected behavior, and ultimately to its impact on the agent's behavior sequence. For example, *conscientiousness_factor* $= 0.1 \longrightarrow$ *organized_trait* $= 0.083 \longrightarrow$ *expressed_behavior* $=$ careless, and ultimately a plan is generated for serving the customer the incorrect item.

2.4. la VIDA

la VIDA is a system to generate value and identity driven autonomous behavior in artificial agents. It is a goal-driven autonomy (GDA) system [21][22] designed according to the goal generation and management schema [2]. As a collection of subsystems interacting to manage and plan for goals, its key functionality is to create an action sequence for an agent that is consistent with its identity profile and can be executed within the current environment.

The identity profile is the key component of *la VIDA*. To the best of our knowledge, an identity profile like the one we have designed hasn't been used to drive agent behavior in a

GDA system. The identity profile is literally the source for the pool of goals that an agent can potentially pursue. It also impacts how actions are planned for and ultimately performed. The identity profile is implemented as a collection of data structures and will be integrated into the agent class as one or more data members. The exact implementation is still being designed, but there are many structures that can represent the graphs and tables that an identity profile is composed of. In future work we will explore the possibility of allowing interaction between an identity profile and a common sense knowledge graph (CSKG) such as ConceptNet [23]. This connection could provide additional context and inference powers to the contents of an identity profile.

3. Identity Profile Representation

3.1. Schema Graphs

Our selected representation and function of roles, values, beliefs, and attitudes are founded on schemata. Each is a graph of related ideas that simulate the organization and functionality of schema. A role describes a function or part played by an agent in social situations. We represent a role as a single word or phrase stored in the network root node where children nodes are strongly correlated verbs, adjectives, and nouns. We name this structure a *role schema* and its primary function is to group together a collection of concepts that goals and actions can be extracted from. In Fig. 1 the role is for "barista", a server and coffee maker within a restaurant specializing in coffee beverages. The noun "server" is identified, as a starting point for identifying actions that a server would typically perform. A CSKG can perform strength of association calculations between any node pair, but additionally can help extract actions and goals from a role schema.

Schemata for values, beliefs, and attitudes have the same function and structure; they are named *value schema*, *belief schema*, and *attitude schema* respectively. Each target concept is placed at the root node and has children that are highly correlated adjectives, and nouns. The agent seeks alignment with its identity and thus values represent the type of behavior the agent strives for. Value schemata hold keywords that the agent would like to be true about itself. For example, a value schema for "hard worker" might have children nodes "persistent", "accurate", "punctual", and "skilled".

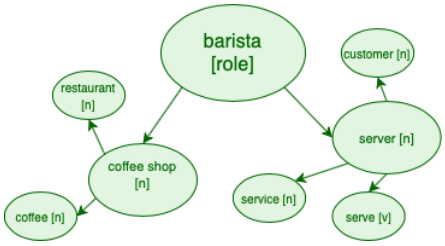

Figure 1: Role Element

A belief schema is a collection of concepts that the agent believes are associated; these concepts may not actually be associated. The purpose of a belief schema is to bias the actions of

an agent in a subtle way. If the agent holds a belief schema for "customer" and it has children nodes "talkative" and "easygoing", a barista agent who has the goal to earn tips, may prioritize socializing with customers over completing their order.

An attitude schema describes the feelings an agent has about a particular concept. This concept is in the root node and its children are the feelings and reactions the agent has in relation to the concept. An attitude schema for "work" might have children "exhaustion" and "boring". Schemata will be manually created at this point of our work, and an agent's identity profile will consist of a small, non-exhaustive collection of schemata for its key roles, values, beliefs, and attitudes. In addition, we will instantiate identity profiles to be compatible with the given scenario.

3.2. Trait-Value Mappings

Mappings between traits and behaviors are essentially functions that return one in a set of behaviors based on the trait strength. The behaviors in the set are similar, only varying by intensity; *extraversion* could be mapped to the behaviors: completely avoid interaction with other agents, interact with agents half of the time, or interact with agents at each opportunity. Trait-value mappings are conceptually represented as tables.

Each agent identity profile has a single personality data member, called a *personality profile*. This key-value pair container holds a character corresponding to each personality factor and it's value. As noted earlier, the value must lie in the closed interval 0 to 1. The lowest strength is denoted with a value of 0 and the greatest strength with a value of 1. For this first iteration of the identity profile we are experimenting with the Mini-Markers PI [11]. This inventory has 40 traits, with a distinct subset of eight traits used to score each of the five factors. Each trait can be attributed a discrete value from 1 to 9; 1 corresponding to "extremely inaccurate" increasing up until 9 as "extremely accurate". All traits have a Varimax-Rotated factor loading for the big five factors within the interval $(-1.05, 1.05)$.

Factor	Value
C	0.9
A	0.7
N	0.81
O	0.32
E	0.92

Trait	Value	Impacted areas:	Weak	Medium	Strong
Energitic	0.9024	body movement, facial movement, dialogue	body movement speed: slow, facial movement speed: slow, dialogue: low verbose	body movement speed: medium, facial movement speed: medium, dialogue: medium verbose	body movement speed: fast, facial movement speed: fast, dialogue: fast verbose

Figure 2: Personality Element and Trait Behavior Mapping

To obtain the score, subset trait ratings are aggregated, each taking the sign of its factor loading with the greatest absolute value. Factor loadings indicate the strength of association between a trait and the FFM factors. For example, the subset of *extraversion* traits are: talkative, extroverted, bold, energetic, shy, quiet, bashful, and withdrawn. It is most strongly positively associated with talkative and negatively with shy. As the inventory is setup, not all FFM factors have the same possible minimum and maximum scores, but the overall minimum and maximum scores are -7 and 7 respectively. Thus, factor scores taken from the test will be normalized by

diving by 7 to be consistent with identity profile 0 to 1 values. Since we are essentially seeking the inverse of the Mini-Markers PI, we use the strength of any factor and factor loadings to infer which traits the agent most strongly expresses. This is not necessarily the correct interpretation and is simply the approach we are experimenting with at present. For example, if an agent has a high *extraversion* value, it will have the traits talkative and extroverted. In addition, if the agent is also high in other factors, a trait's factor loadings may combine to attribute the trait to the agent; a trait should have a value $> |threshold|$. Continuing with the example, if the agent also has a high *neuroticism* value, factor loadings could combine to attribute the bold trait to the agent. Trait strength is calculated by multiplying the factor strength by the trait factor loading. Trait intensities are quantized into three categories; weak, medium, and strong. Values in [0,0.33) are weak, values in [0.33,0.66] are medium, and values in (0.66,1] are strong.

	Factor				
Item	**I**	**II**	**III**	**IV**	**V**
Talkative	**.73***	.14	-.12	-.05	-.05
Extroverted	**.70***	.07	-.07	.11	-.01
Bold	**.51***	-.17	.00	.24	.03
Energetic	**.44***	.18	.18	.18	.02
Shy	**-.79***	.15	.04	-.08	-.03
Quiet	**-.76***	.02	.13	.05	.07
Bashful	**-.73***	.19	.04	-.06	-.06
Withdrawn	**-.71***	-.15	-.07	-.10	.02

Figure 3: Mini Markers Extroversion Subset [11]

4. Conclusions

In this work we explore some early ideas for the representation and function of an agent identity profile. The identity profile captures the mental aspects that describe an agent, such as its roles, values, beliefs, and attitudes. We also consider how the elements of the agent's identity profile motivate it and what type of behavior results from those motives. As this is early work, we have many ideas of how it can be further developed. We will begin with implementation and testing, as this will reveal many design short-comings and necessities; redesign will follow. The *trait-value mapping* will be extended to cover the three intensities of behavior for each of the 40 Mini-Marker traits. To complete the table, we must determine which aspects of agent behavior should be impacted by a particular trait and in which manner. In Fig. 2 we use the trait "energetic" as an example, and note that the body movements, facial movements, and dialogue should be impacted by this trait. Similar decisions will need to be made for each of the remaining traits.

References

[1] R. M. Ryan, E. L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, American Psychologist 55 (2000) 68–78.

[2] M. Hanheide, A framework for goal generation and management, Proceedings of the AAAI Workshop on Goal-Directed Autonomy (2010).

[3] J. L. Wyatt, A. Aydemir, M. Brenner, M. Hanheide, N. Hawes, P. Jensfelt, M. Kristan, G.-J. M. Kruijff, P. Lison, A. Pronobis, K. Sjöö, A. Vrečko, H. Zender, M. Zillich, D. Skočaj, Self-understanding and self-extension: A systems and representational approach, IEEE Transactions on Autonomous Mental Development 2 (2010) 282–303.

[4] S. M. Sincero, Motivation and emotion, 2012. URL: "https://explorable.com/motivation-and-emotion".

[5] S. Reiss, Multifaceted nature of intrinsic motivation: The theory of 16 basic desires, Review of General Psychology 8 (2004) 179–193. doi:10.1037/1089-2680.8.3.179.

[6] P. Howard, J. Howard, The big five quickstart: an introduction to the five-factor model of personality for human resource professionals (1995).

[7] A. Lieblich, R. Josselson, Identity and narrative as root metaphors of personhood, 2009, pp. 203–222. doi:10.1017/CBO9781139086493.015.

[8] R. Pereira Ramos, R. Eduardo da Silva, J. C. Koerber Reis, A personality model based on reiss motivational profile for autonomous digital actors (2012).

[9] E. Cebolledo, O. Troyer, iattac: A system for autonomous agents and dynamic social interactions – the architecture, 2015, pp. 135–146. doi:10.1007/978-3-319-19126-3_12.

[10] R. Sun, Motivational representations within a computational cognitive architecture, Cogn. Comput. 1 (2009) 91–103.

[11] G. Saucier, Mini-markers: A brief version of goldberg's unipolar big-five markers, Journal of Personality Assessment 63 (1994) 506–516.

[12] J. Marcia, Development and validation of ego-identity status, Journal of Personality and Social Psychology 3 (1966) 551–558.

[13] R. B. Evans, William james, "the principles of psychology," and experimental psychology, The American Journal of Psychology 103 (1990) 433–447.

[14] R. C. Schank, Tell me a story : a new look at real and artificial memory, Scribner, New York, 1990.

[15] P. Winston, The right way, Advances in Cognitive Systems 1 (2012).

[16] H. Markus, Self-schemata and processing information about the self, Journal of personality and social psychology 35 (1977) 63–78.

[17] G. H. Bower, J. B. Black, T. J. Turner, Scripts in memory for text, Cognitive Psychology 11 (1979) 177–220.

[18] R. C. Nemanick, D. C. Munz, Extraversion and neuroticism, trait mood, and state affect: A hierarchical relationship?, Journal of social behavior and personality 12 (1997) 1079–.

[19] D. S. Chiaburu, I.-S. Oh, C. M. Berry, N. Li, R. G. Gardner, The five-factor model of personality traits and organizational citizenship behaviors: A meta-analysis, Journal of applied psychology 96 (2011) 1140–1166.

[20] A. A. Uliaszek, R. E. Zinbarg, S. Mineka, M. G. Craske, J. M. Sutton, J. W. Griffith, R. Rose,

A. Waters, C. Hammen, The role of neuroticism and extraversion in the stress-anxiety and stress-depression relationships, Anxiety, stress, and coping 23 (2010) 363–381.
[21] H. Muñoz-Avila, Adaptive goal driven autonomy, Case-Based Reasoning Research and Development. ICCBR 11156 (2018) 3–12.
[22] M. E. Klenk, Goal-driven autonomy in planning and acting, 2010.
[23] R. Speer, J. Chin, C. Havasi, Conceptnet 5.5: An open multilingual graph of general knowledge, Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (2017) 4444–4451.