SITARA SYSTEMS · TECHNICAL REPORT

Modelling Autonomous Agent Psychology

Using Personality Traits

Nathan S. Lachenmyer and Sadiya S. Akasha · Sitara Systems

Technical Report · 2026

Abstract

Autonomous agents are a standard tool for animating groups, but the groups they produce are usually homogeneous: every agent runs the same behavioural model, so individual differences come from an agent's momentary situation rather than from any stable disposition. Giving agents durable, distinct characters normally means hand-tuning each one's low-level parameters, which is slow and rarely yields a coherent temperament, because the parameters that must move together to read as a single character are not obvious in advance. A known remedy is to drive those parameters from a personality model, so that a small, legible description stands in for dozens of settings; this has been used to add variation to simulated crowds. This report describes what happens when the same idea is carried out of the screen and into a physically realised setting: we used the five-factor (OCEAN) personality model to individuate the behaviour of twelve autonomous robots in a public art installation. We set out how each behavioural parameter is coupled to one or more personality axes, why the coupling — not the axes themselves — is where the expressiveness lives, and what we observed when visitors met the result over a months-long museum run. The aim is less to introduce a technique than to report, from practice, how well a personality-parameterised behavioural model holds up when the agents are real, embodied, and shared with the public.

1 Introduction

People readily attribute inner life to things that move. Shown even a very simple behaviour, an observer will assign it an emotion(Braitenberg, 1984); shown that behaviour repeatedly, the observer begins to speak of the agent as having a personality — a stable pattern of temperament and response that carries from one encounter to the next. Homogeneous models of group motion, such as flocks of birds(Reynolds, 1987) or schools of fish(Stephens et al., 2003), leave room for an agent to act individually in a given moment, but because every agent shares the same underlying model, none of them carries a personality forward across interactions. An agent that strikes an observer as aggressive in one moment is not, in any lasting sense, more aggressive than its neighbours; the impression does not survive the next encounter.

Part of the difficulty is that building a believable individual mind is laborious. Even with ethological studies to draw on, giving an artificial organism life-like motion takes a great deal of manual tuning, and a single behavioural model may expose dozens of parameters. Finding the particular combination of parameters that shifts an agent's long-term temperament in a coherent way — rather than merely changing one behaviour in isolation — is genuinely hard, and it falls to the animator, not the model, to do it.

The response we adopt is not new in outline. A designer can choose an agent's high-level disposition in the vocabulary of personality and let the low-level parameters be derived from that choice; this idea has an established history in crowd simulation, reviewed in the next section. What this report contributes is not the mechanism but the transfer: we take a personality-parameterised behavioural model out of screen-based crowd work and into a fleet of physically realised agents in a public setting, and report how it behaved there.

2 Background and related work

The intuition this work draws on is old. Even simple motion invites an observer to read in emotion and, over repeated encounters, personality(Braitenberg, 1984). The believable-agents tradition made that intuition a design goal, treating a legible inner life — rather than physical or behavioural fidelity — as the thing worth authoring(Bates, 1994).

The specific move we use — mapping the five-factor personality model onto the low-level parameters of a behavioural model — was worked out in crowd simulation. Durupınar and colleagues tied each OCEAN factor to several parameters of a crowd simulator, letting an author set a crowd's behaviour from five personality dimensions instead of a dozen raw controls, with some factors coupled to a parameter positively and others inversely(Durupınar et al., 2008, 2011). Their goal was aggregate realism: a crowd that reads as heterogeneous, validated through perception studies of rendered footage. The mechanism we use is essentially theirs, and the practice of coupling several parameters to a single axis — including deliberate inverse couplings — originates in that work.

Two things separate the present report from that lineage. The setting is physical rather than simulated: the agents are embodied, non-anthropomorphic robots with real sensing, deployed in public. Most personality work in human–robot interaction concerns how a robot's personality is *perceived*, or how it might be matched to a user, rather than how five personality dimensions drive the low-level motion of a fleet(Robert et al., 2020); that is the gap this deployment sits in. And the goal is different: not a statistically varied crowd but a small set of agents each meant to be recognised as a specific, persistent individual. That shift — from aggregate variation to legible individuation — is what makes the coupling logic load-bearing, and it is what the rest of this report examines.

3 Design

In cognitive and personality psychology, the five-factor model — commonly abbreviated OCEAN — is the dominant account of how general personality is structured(Widiger and Crego, 2019). Its five factors are treated as independent, or orthogonal, dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism. Openness is a receptiveness to new experience, whether aesthetic, cultural, or intellectual; conscientiousness is the tendency to be organised, responsible, and hardworking; extraversion is the orientation of energy outward, toward the external world rather than the inner one; agreeableness is a concern for social harmony; and neuroticism is a tendency to dwell on negative emotions and outcomes.

The model has been reproduced across many different populations(Widiger and Crego, 2019) and has even been extended to describe personality in animals(Vékony et al., 2022), which makes it a reasonable organising structure for agents that are neither human nor, strictly speaking, animal. To keep the framework general, we treat each trait as a bundle of the descriptive adjectives in Table 1, and let those adjectives stand in for the trait when we reason about behaviour.

Trait	**Descriptive Adjectives**
Openness	curious, risk-taking, responsive
Conscientiousness	focused, diligent, persistent
Extraversion	social, energetic, assertive
Agreeableness	communal, tolerant, patient
Neuroticism	avoidant, defensive, cautious

Table 1. Descriptive adjectives associated with each of the five factors of the OCEAN personality model.

Following this approach, we place a parameterisation layer over the behavioural model: a mapping from the five personality axes to the model's low-level parameters. Each OCEAN factor is treated as a bipolar axis, running from one pole of its adjectives — which we label $+1$, fully characteristic of the trait — through a neutral midpoint to the other, -1, its opposite. A designer places an agent somewhere on each of the five axes; that setting is the agent's personality, and every behavioural parameter is read off from it rather than dialled in by hand.

The model lies not in the axes but in the mapping: in which parameters each axis controls, and in which direction. Two properties of that mapping do most of the work. First, a single axis can drive several parameters at once, so that moving one dial shifts a whole cluster of behaviours together rather than one in isolation. Second, parameters can be coupled to an axis in opposing directions — tightening one while loosening another — and parameters that a person would expect to trade off can be tied to opposite poles of the same axis. It is this coupling of related parameters to shared, or deliberately opposed, psychometric axes that makes an agent read as a coherent temperament rather than a collection of unrelated settings, and that lets a single, human-meaningful decision — how open or how neurotic an agent should be — move the underlying numbers together, in the direction intuition expects. Table 2 sets out the couplings we used.

4 Implementation

We applied this technique to the agents of an interactive installation: a group of twelve lighter-than-air robots that could sense and respond to the people around them. The installation and the artificial-life system beneath it — Anicka Yi's *In Love With the World* (Hyundai Commission, Tate Modern, 2021) — are described in full in a companion paper(Lachenmyer and Akasha, 2022a,b); here we focus on how personality traits gave the agents distinct and durable individual behaviour. Our aim was for visitors to recognise each agent as having its own way of behaving — its own "personality" — rather than seeing twelve interchangeable copies.

Each agent, an Aerobe, carried a thermal camera that let it pick out the warm signatures of nearby people. Its behavioural repertoire was small but expressive: it could seek out a person, hover and bob up and down above someone it had found, gather with other Aerobes into a slow circling flock overhead, and dart away from anything it treated as a threat. Each of these behaviours was governed by a few parameters. In seeking, an Aerobe had a minimum radius it would search within (r_{min}) and a distance over which it could "see" a thermal signature (d_{sight}). In bobbing, it had a maximum height it would rise to (h_{max}) and a length of time it would continue (t_{bob}). When several Aerobes converged on the same visitor, each had its own threshold n for how many others had to be present before it would join the flock, and its own time t_{flock} after which it would drift away. Darting was set by a distance from the threat at which the agent would flee (d_{dart}) and a time it would keep fleeing (t_{dart}).

Behavior	Parameter	Trait Mapping
Seek Visitors	r_{min}	O−
	d_{sight}	O+
Bob Up and Down	h_{max}	E+
	t_{bob}	C+
Overhead Flocking	n	E−
	t_{flock}	A+
Dart	d_{dart}	N+
	t_{dart}	N+

Table 2. Behavioural parameters and their trait mappings. Traits are identified by the first letter of their name (O, C, E, A, N); a $+$ indicates that the parameter and trait are positively correlated, and a $-$ that they are negatively correlated.

Table 2 sets out these couplings concretely, and several of them tie more than one parameter to a single axis. Openness governs both how far an Aerobe can see (d_{sight}, which rises with it) and the minimum radius within which it bothers to look (r_{min}, which falls), so that a more open agent both reaches farther and attends more closely — two facets of the same curiosity, moved by one dial. Extraversion raises how high an agent bobs and lowers the number of companions it needs before joining a flock; neuroticism governs both how far an agent stays from a threat and how long it flees. In each case the parameters coupled to an axis are ones a person would expect to move together, which is what lets a single decision about personality produce behaviour that hangs together rather than a set of independent adjustments.

Figure 1. An Aerobe interacting with a crowd of visitors underneath. (© Sitara Systems)

5 Observations in deployment

The installation ran for several months at Tate Modern, a long and uncontrolled exposure to a general public — the opposite of a lab study, and a demanding test of whether individuated behaviour reads at all outside a controlled setting. Our observations are informal, drawn from watching visitors and from unstructured conversation with them rather than from an instrumented study, and should be read as such.

What we saw was consistent. Visitors readily assigned distinct personalities to the Aerobes, and described individual agents as having a character that held within a session and was recognisable from one session to the next. People singled out particular agents, followed them around the hall, and speculated about what a given Aerobe *wanted* — treating its behaviour as evidence of motive rather than of mechanism. That speculation opened richer conversation — about how the agents worked, and about what the piece was for — than we had seen when agents behaved interchangeably. For a work whose aim is to make people curious about a machine's inner life, a legible personality proved a direct route to that curiosity.

We report this as encouragement, not proof. The evidence is qualitative and unblinded, and the installation was charismatic in ways that go well beyond the personality model. But the effect the crowd-simulation studies established under controlled viewing — that observers attribute traits to parameterised motion — appears to survive the move to embodied agents and an unprompted public, which is the claim this deployment is really in a position to support.

6 Discussion

Compared with tuning low-level parameters directly, this lets a behavioural designer work in terms of higher-level ideas that already match how people perceive character. Mapping parameters to traits brings two advantages. First, the designer no longer has to understand the physics or the computational machinery that produces a behaviour. Second, the space they work in shrinks from dozens of loosely related numbers to five parameters. Because those five correspond to familiar dimensions of personality, they are easy to reason about, and they keep the visitor's experience of the agent — rather than the mechanics of the simulation — at the centre of each design decision.

None of this is a new mechanism, and we do not claim it as one; the value here is in the choice and the setting. It is worth being explicit about why a deterministic five-dial mapping is the right tool for this job, given that personality is now often induced in agents by much heavier means. A physically embodied installation that must run untended for months rewards a control layer that is inspectable, cheap, and predictable: a designer can reason about exactly how a personality setting will move the machine, and can trust it to behave the same way in month four as on opening night. A learned or generative policy would buy flexibility this work does not need, at a cost in legibility and operational life it cannot afford. The method earns its place not by being the most powerful option but by being the most maintainable one that meets the requirement.

That trade sits at the centre of how Sitara approaches behavioural modelling for spatial human-computer interaction (HCI) — the design of computational systems at the scale of rooms and buildings rather than screens. Installations of this kind are long-lived public infrastructure as much as they are artworks, and the methods that sustain them are chosen for legibility and operational life, not novelty. Individuating a fleet of agents through a small, legible personality model is a modest instance of that larger practice: a technique borrowed intact from adjacent work, and made to earn its keep in the built environment.

References

Joseph Bates. The Role of Emotion in Believable Agents. *Communications of the ACM*, 37(7): 122–125, 1994. doi: 10.1145/176789.176803.

Valentino Braitenberg. *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1984. ISBN 9780262521123.

Funda Durupınar, Jan Allbeck, Nuria Pelechano, and Norman Badler. Creating Crowd Variation with the OCEAN Personality Model. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

Funda Durupınar, Nuria Pelechano, Jan Allbeck, Uğur Güdükbay, and Norman I. Badler. How the OCEAN Personality Model Affects the Perception of Crowds. *IEEE Computer Graphics and Applications*, 31(3):22–31, 2011. doi: 10.1109/MCG.2009.105.

Nathan S. Lachenmyer and Sadiya S. Akasha. An Aquarium of Machines: A Physically Realized

Artificial Life Simulation. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 5(4), 2022a. doi: 10.1145/3533388. Describes Anicka Yi's *In Love With the World* (Hyundai Commission, Tate Modern, 2021) and the artificial-life system driving the Aerobes.

Nathan S. Lachenmyer and Sadiya S. Akasha. Creating Life-like Autonomous Agents for Real-time Interactive Installations. In *ACM SIGGRAPH 2022 Talks*. Association for Computing Machinery, 2022b. doi: 10.1145/3532836.3536255.

Craig W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1987*, pages 25–34. Association for Computing Machinery, Inc, aug 1987. ISBN 0897912276. doi: 10.1145/37401.37406.

Lionel P. Robert, Rasha Alahmad, Connor Esterwood, Sangmi Kim, Sangseok You, and Qiaoning Zhang. A Review of Personality in Human–Robot Interactions. *Foundations and Trends in Information Systems*, 4(2):107–212, 2020. Also available as arXiv:2001.11777.

Kingsley Stephens, Binh Pham, and Aster Wardhani. Modelling fish behaviour. In *Proceedings of the 1st International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, GRAPHITE '03*, pages 71–78, 2003. ISBN 1581135785. doi: 10.1145/604471.604488.

Kata Vékony, Fruzsina Prónik, and Péter Pongrácz. Personalized dominance – a questionnaire-based analysis of the associations among personality traits and social rank of companion dogs. *Applied Animal Behaviour Science*, 247:105544, feb 2022. ISSN 0168-1591. doi: 10.1016/J.APPLANIM.2022.105544.

Thomas A. Widiger and Cristina Crego. The Five Factor Model of personality structure: an update. *World Psychiatry*, 18(3):271, oct 2019. ISSN 20515545. doi: 10.1002/WPS.20658. URL /pmc/articles/PMC6732674/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6732674/.