Designing Naturalistic Simulations for Evolving AGI Species

Christian Hahm
Department of Computer & Information Sciences, Temple University, Philadelphia, PA, U.S.A.

Keywords: Adaptation, Artificial Evolution, Genetic Algorithm, Virtual Environment, Artificial Life, AI Embodiment, Evolutionary Simulation.

Abstract: This paper identifies basic principles for designing and creating evolutionary simulations in the context of general-purpose AI (AGI). It is argued that evolutionary simulations which employ certain nature-inspired principles can be used to evolve increasingly intelligent AGI species. AGI frameworks are particularly suited for evolutionary experiments involving embodiment since they can operate arbitrary evolved bodies. Once a designer manually defines a simulation's initial conditions, each run is an automated exploration of a novel subset of species. In this way, naturalistic simulations generate huge amounts of empirical data for evaluating the robustness of AGI frameworks, along with many promising species that can be later instantiated in other simulated environments or even physical robots for practical applications.

1 INTRODUCTION

Virtual simulation goes naturally with AI: since both are implemented in digital computers, it is almost trivial to interface the two. We should consider this a huge benefit, since virtual environments (which, for this discussion, can be referred to using other words e.g., a "world", "universe", "reality", "simulation" etc.) are highly customizable by the designer. To perform "simulation" in the context of AI means:

1.) To design and create a *virtual* world (in a computer, using a "simulator" software), and

2.) To *embody* one or more AI agents *in that world*.

The purpose of simulation is to provide AI agents a controlled environment in which to interact and learn from their simulated experience, such that we may study the agents (e.g., to observe their processes under certain conditions) or utilize them in some real-world application once they are sufficiently trained.

An agent is considered "embodied" in an environment so long as it has sensorimotor mechanisms (Wang, 2009), meaning *sensors* (i.e., tools on the body which transduce environmental signals as inputs *to* the mind) and *actuators* (i.e., tools on the body controlled by output signals *from* the mind) in that environment. The agent receives sensory signals to monitor the environment, and can send output signals to move its body. Like a "brain-in-a-vat", an AI system treats sensory signals identically regardless of whether they come from simulated virtual sensors or real-world physical sensors.

Artificial general intelligence (AGI) is a term which can be used to delineate cognitive-inspired computer systems with full autonomy and general-purpose ability. The possible applications of simulation to AGI research are numerous, but mostly involve optimizing the system on either the object level or the meta level. In object-level testing, we embody some agents and let them learn simulated tasks autonomously. In meta-level testing, we vary configurable *cognitive parameters* to test various "personalities" of AGI individuals on object-level tests. Cognitive parameters may range from scalar control values (e.g., decision thresholds, attention decay, etc.) to structures in the system's architecture (e.g., buffers, memory, etc.). The exact nature of cognitive parameters depends on the selected framework.

Since AGI systems are general-purpose, they should be capable of operating arbitrary bodies. In considering embodiment, we can expand the notion of parameters from cognitive values to include physical structures like the system's morphology, sensors, actuators, and physiology (*bodily parameters*). Any parts of the system can be parameterized, except those core aspects which we desire to remain static. When aspects of the system's mind and body are parameterized, it evokes the possibility of creating varied AI "species": agents which share a common core framework of intelligence, but vary by their exact cognitive

Hahm, C.
Designing Naturalistic Simulations for Evolving AGI Species.
DOI: 10.5220/0011299600003274
In *Proceedings of the 12th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2022)*, pages 296-303
ISBN: 978-989-758-578-4; ISSN: 2184-2841

processes, bodily form, and sensorimotor capabilities.

Evolutionary simulations automate the process of finding and optimizing such mind-body parameterizations. Traditionally, evolutionary simulations propagate species that maximize a quantitative "fitness" function. Natural evolution admits no such fitness function, and to simulate it requires a slightly different treatment than traditional genetic algorithms. Instead of optimizing genomes to maximize a predefined fitness function, "naturalistic simulations" require agents to prove their reproductive fitness independently, encouraging the evolution of intelligent autonomy. If the selected AGI framework is flexible and the simulated environment is sufficiently rich, the resulting species should exhibit interesting adaptive behaviors.

2 RELATED WORKS

Evolving autonomous machines is not a novel concept; one of the earliest and most famous attempts at elaborating this idea is mathematician John von Neumann's self-reproducing automaton. He hypothesized a baseline level of complexity that allows the evolution of increasingly complex systems (von Neumann, 1966, p.78-80), realized by the three sub-processes of reproduction and evolution: *duplication*, *modification*, and *instantiation* of genomes. In his context, the automaton system itself performs all three subprocesses, including modifying the genome.

Though "AGI" is often construed as human-level intelligence, our simulated agents will be more like primitive animals than humans. They will not be capable of manually modifying their own genes; the task of genomic modification belongs to the simulation. As long as the agents handle reproduction autonomously, the simulation design will handle the actual evolutionary processes behind the scenes. Game engines such as *Unity3D* (Juliani et al., 2018) are the ideal simulator software since they facilitate building simulated physical worlds and arbitrary scripting.

(Holland, 1992) formalizes a methodology for approaching so-called "problems of adaptation". This work led to the popularization of *genetic algorithms*, an evolutionary computing method by which to explore a space of evolving structures using conceptions of fitness, genetics, and reproduction. Genetic algorithms are arguably the best approach for simulating natural evolution, because they implement the general principles that make natural genetics adaptive. Namely, genetic algorithms use "survival of the fittest", recombination, and random mutation to explore a certain space of evolving structures in search of more optimal (i.e., "fit") structures.

(Sims, 1994a; Sims, 1994b) use Holland's genetic algorithms to evolve various agent bodies and their neural networks, including by facing the agents off in direct physical competition (Sims, 1994a). The agents are encoded using a highly compact and flexible genetic language in the form of directed network graphs. Physical body parts contain sensors (e.g., contact sensor, light sensor) or actuators (e.g., rotational force on joints), and each body part has a neural circuit allowing for some level of distributed control.

(Soros and Stanley, 2014) propose four necessary conditions for open-ended artificial evolution: 1&2.) new and existing individuals must be required to meet some non-trivial minimum criterion (MC) before they can reproduce, so species do not degenerate into trivial-behavior automatons, 3.) individuals must autonomously meet that MC, 4.) the complexity of the individual's genotype can grow unrestricted. The naturalistic MC is the ability to reproduce, including the requisite apparatus and motivation (Soros and Stanley, 2014, p.2-3).

(Strannegård et al., 2020) establishes a "generic animats" framework in which simulated organisms are defined by homeostatic variables, sensors, patterns, motor, actions, and reflexes. (Strannegård et al., 2021) formalizes ecosystems containing *inanimate objects* and *organisms* that have unique properties (e.g., digital "animals" have energy levels, environmental objects have certain chemical reactivity) and common properties (e.g., all objects have mass).

For this discussion we will assume Wang's working definition of *intelligence*: "*the capacity of a system to adapt to its environment with insufficient knowledge and resources*" (Wang, 1995; Wang, 2019, p.13; p.17-18). According to this definition, the minimum requirement for intelligence is to operate under an Assumption of Insufficient Knowledge and Resources (AIKR). A system operating under AIKR: 1.) uses *constant* computational resources, 2.) works in *real-time*, and 3.) is *open* to new tasks regardless of their content. In plainer terms, such a system is like a living organism: it is a fast, finite agent that operates effectively even under uncertainty, while constantly learning from its experience so as to better achieve its goals. Cognitive frameworks designed with AIKR in mind may be flexible enough to permit the evolution of complex adaptive behaviors.

3 SIMULATION DESIGN

3.1 Formalizing the Problem

Following (Holland, 1992, p.28,35), a given *problem of adaptation* and the *adaptive system* to tackle it, in our case evolving AI species in simulation, are defined by the variables: $\mathcal{A}, \Omega, \mathcal{T}, \mathcal{E}, \chi$.

The variable $\mathcal{A} = \{A_1, A_2, ...\}$ is the *set of all genomes* that can possibly be evolved; if the simulation's genetic language permits genomes to grow without restriction, this set is infinite. An evolutionary run searches this set. Genome encoding should include both *cognitive parameters* for the AGI system and *bodily parameters* for its embodiment.

$\Omega = \{\omega_1, \omega_2, ...\}$ is the *set of genetic operators* that can be used to modify genomes. A given operator might perform some variation of crossover or mutation.

$\mathcal{T} = \{\tau_1, \tau_2, ...\}$ is the *set of possible reproductive plans*, each of which is a sequence of operators from Ω which can be used to incrementally traverse $\mathcal{A}$. One τ may be selected for the entire simulation run, or the choice of τ could even be evolved.

$\mathcal{E} = \{E_1, E_2, ...\}$ is the *set of possible environments*. Creating a concrete simulation requires specifying the actual environment $E \in \mathcal{E}$.

Finally, χ represents a *criterion* to compare the many possible plans in $\mathcal{T}$.

So, to specify these variables is to determine the possibilities of the evolutionary simulation. The contents of plans in $\mathcal{T}$ depend on the operators available in Ω; for example, a plan could look like $\tau = (\omega_i, \omega_j, \omega_i, \omega_k, ...)$. Operators in Ω depend on how the designer specifies the genetic language for $\mathcal{A}$. Therefore, this discussion will mostly ignore the dependent variables Ω and $\mathcal{T}$, instead focusing on general considerations for specifying $\mathcal{A}$, E, and χ.

3.2 Embodiment

3.2.1 Atoms

One major design problem of virtual simulation is that we have to manually specify its irreducible components. There are no interacting molecules like there are in reality, unless we program them. We want to simulate a physical reality efficiently and with finite resources. We also want a flexible simulation, which requires irreducible generic components that can combine and interact in various ways.

The simulation can implement a notion of *equal exchange*, where a finite number of virtual *atoms* underlie all structures in the simulation. There could be

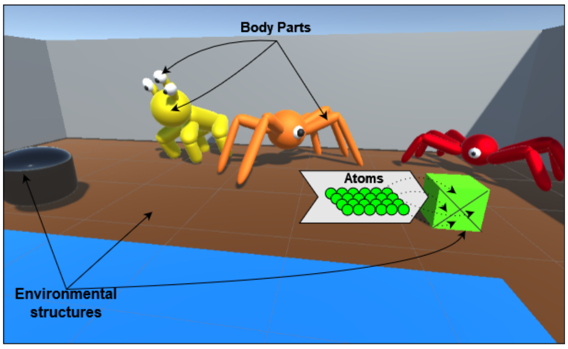

Figure 1: *Atoms* are the underlying virtual representation of a simulation's structures. In this mockup, the green block is an environmental structure consisting of atoms. Agents could consume the atoms to exchange them for homeostatic maintenance, physical growth, or offspring.

various types of atoms with unique properties. The atoms are abstract intermediary components, never explicitly seen but used justify *environmental structures* in E and physical expression of *body parts* encoded by $\mathcal{A}$ (see *Figure 1*). Atoms could correspond one-to-one with size-standardized polygons or voxels. This prevents any structure from growing disproportionately compared to the others by enforcing equal exchange.

While pure morphology can be grown incrementally, polygon by polygon, some body parts must be hardcoded with special-purpose functionality in the form of modality-specific sensors, actuators, and motivation. For example, vision requires a special photoreceptor body part which can render a partial viewpoint of the scene and communicate it to the AGI system. One way to handle this is to exchange atoms one-to-one for pure structure (with some distinguishing properties) whereas require a many-to-one exchange for structures with special functionality. For a simple example, two red and two green atoms could correspond to a single photoreceptor, meaning a parental agent will need to consume those atoms to instantiate an offspring with a photoreceptor part. Agents would return their body's absorbed atoms to E upon death (via simulated decomposition) for recycling.

The minimum E is an environment that hosts reproducing (thus evolving) cognitive agents and respects a conservation of materials. Once this is achieved, E can be made more complex and interactive using frameworks like atoms, perhaps even including genomes for other kingdoms of life (e.g., plants, fungus, etc.). Various "organisms" could reorganize atoms in unique ways, in a simulated "circle of life". Non-cognitive entities like plants are not strictly necessary, but without them the environment will be simplistic and should not be expected to produce very sophisticated or interesting agents.

3.2.2 Sensorimotor

In order to evolve sensorimotor, first the environment E must simulate specific modalities (e.g., floating odorants, sound waves), and the genomes in α should be capable of expressing the corresponding modality-specific sensors and actuators (e.g., olfactory organs, visual organs, vocal cords, etc.). It is certainly worth exploring ways to evolve arbitrary sensory modalities. However, since sensors measure physical signals (e.g., goodness of molecular fit, mechanical pressure, wavelength of light, etc.) which do not exist by default in simulation, it seems necessary to explicitly support certain modalities in the design of E and α.

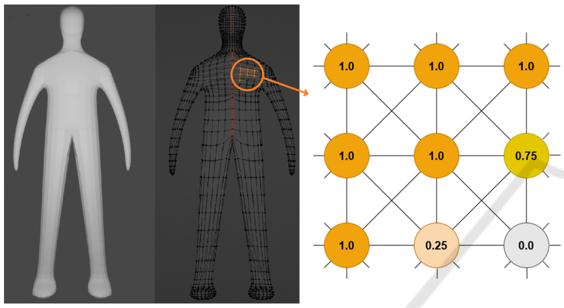

Figure 2: Sensors such as tactile sensors on a 3D model (left, middle) should be represented "topographically", where physically adjacent sensors are also represented adjacently, such as in a network (right). (Made with Blender, blender.org).

Natural sensorimotor neurons in the body bundle then connect to the brain in spatially ordered "topographic" mappings (see (Wolfe et al., 2006, p.64,280,397)). This helps the agent understand the spatial distribution of its sensations. Simulated sensorimotor signals could be represented in a similar way, where sensors of the same type record their sensations in a topographic map. The map should be "topographic" in that sensors which are physically close and adjacent in the simulation are also represented adjacently in the map (see *Figure 2*).

For example, it is quite intuitive to simulate simple vision: render a full 2D visual image of the scene, and treat each pixel of the image as a single visual sensor (aka photoreceptor). The situation is muddier for the other senses. More generally, topography can be described using a network, where edges explicitly (as in a graph, where nearby nodes are connected) or implicitly (as in a convolutional neural network, where neighboring nodes connect to a common node) represent adjacency. A network representation might be useful for touch perception, which relies on unevenly distributed sensors and may not be as amenable as vision to a 2D array representation. The AGI system can then use the topography to selectively group and process sensations (Wang et al., 2022).

3.3 Fitness and Reproduction

Traditionally with genetic algorithms there is an additional variable, μ, specifying the function by which to measure an individual's *fitness*. The simulation proceeds in discrete rounds, where after each round the fitness of individuals is quantitatively measured using μ. Then, the genetic material of a few individuals with the highest scores are combined to create many offspring, whereas those with lower scores are culled. In AI research, we would ideally select a μ that measures and thus optimizes for intelligence. However, no widely-agreed μ exists to quantitatively measure "intelligence", amongst humans nor across species. Besides, evaluating agent performance with a single number is inflexible, since such evaluations are more vulnerable to "hacky solutions" and much generality is lost as species evolve towards optimizing only one function.

In nature, "reproductive fitness" is a tautology. There is no numeric fitness measure by which nature decides reproduction, instead fitness *is equivalent to* (or proven by) successful reproduction. Traits which confer reproductive and survival advantages in the current environment will tend to appear more frequently in the populations than traits which provide relatively less advantage. Therefore, a naturalistic simulation with an implicit μ should at the very least yield increasingly prolific species. Reproduction, as in nature, can be done sexually or asexually. Asexual reproduction is simpler to simulate and very quickly grows the population, though agents who are selectively sexual might induce faster speciation and evolved capabilities, as argued in (Canino-Koning et al., 2017).

Although such a μ frees us from needing to explicitly measure fitness, it introduces a number of problems. First, all simulated agents need to meet the natural "minimum criteria" (Soros and Stanley, 2014): they must be capable of reproducing autonomously. "Capable" not only means physically able to perform a reproductive action (as determined by the specific simulation), but also requires the agents to be motivated towards that action (i.e., to have a sex drive). In biological contexts, bodily glands release hormones that modulate reproductive motivation in the brain (Wise, 1987; Fisher et al., 2006; Cummings and Becker, 2012). A bodily origin of motivations in natural agents is interesting to notice in our context, since it implies motivation evolves with the body.

There needs to be a tradeoff between consuming atoms from the environment (e.g., food) and producing offspring. Consistent and balanced resource exchange is critical as it requires agents to earn their success. The exact process of this tradeoff during reproduction depends on the simulation design and α. One could simulate naturalistic processes like gestation (where a parent gets pregnant then gradually consumes the relevant atoms to grow a baby), incubation, external egg fertilization, etc., though the biological details are not necessary to replicate, only an energy/materials tradeoff to instantiate and grow the offspring.

The choice of adaptive plan $\tau \in \mathcal{T}$ for a given simulation run is important, since it determines the evolutionary trajectories through α. Yet, we do not know the best τ to choose. χ is the quantitative criterion for evaluating plans in $\mathcal{T}$ so as to find the best τ. The χ is normally something like average fitness of the population, $\bar{\mu}$, so as to try various τ in application and pick one which maximizes $\bar{\mu}$, but since a naturalistic μ is implicit, we cannot pick such a χ.

One decent option for χ is the total number of individuals produced by τ, which should result in simulations with many agents and hopefully more opportunities for beneficial evolution. However, any χ or τ can be tried — more agents by time t does not necessarily mean "more-intelligent" agents by time t. It may be possible to treat χ implicitly like μ, leaving the choice of τ to evolution by allowing the selection of τ to vary depending on the species (in which case, each $A \in \alpha$ should represent its selection of τ).

3.4 Homeostasis as Seed Motivation

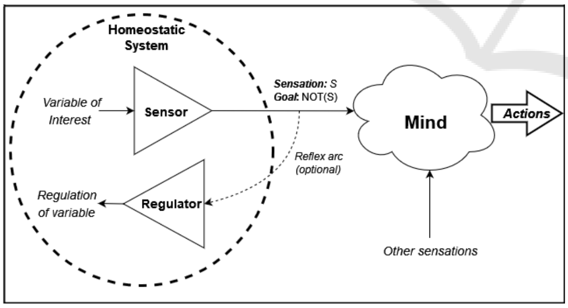

Figure 3: Bodily systems entering homeostatic imbalance may trigger sensations and motivation in the mind.

Homeostasis is the tendency of a system to maintain bodily variables within an optimal range. The purpose of homeostasis is to keep the system alive; whenever there is a homeostatic imbalance, the body triggers automatic reflexes and signals motivation to the mind so as to fix the imbalance. (Tsitolovsky, 2015, p.3,4) explains: "*Needs are at the heart of motivations*", and since homeostatic goals are endogenous (originating from within the agent's own mind-body system, rather than from an external user) they "*turn [the agent] into a subject with its own behavior*". Reflexes only have very limited power to restore the body to stability, depending on the severity and complexity of the issue. This is why, in nature, both sensation *and* motivation are signalled to the central nervous system, so that the agent itself can restore homeostasis by interacting with the environment (Craig, 2003; Tsitolovsky, 2015).

Maintenance of the body as the agent's source of motivation makes perfect sense, as it keeps the agent alive. Determining how agents source motivation is extremely important since without motivation there is no action. A small set of motivational signals from the body is enough. Simple motivational "seeds" can derive a wide range of additional motivations, causing many complex behaviors to emerge depending on the specific agent's knowledge and capabilities (Hahm et al., 2021). Evolving homeostatic systems on the body can provide various seed motivation signals to the agent. The more skilled the agent is at meeting useful homeostatic survival needs, the more likely the agent will survive to reproduce.

One could simulate homeostatic motivation as in *Figure 3*, for example, within a body part containing variable-specific sensors which activate whenever some variable (e.g., heat, cold, pain) veers too far from its acceptable range. The sensor then signals both a sensation of the imbalance $S = Belief(sensation)$ and a negative desire to alleviate it $D = Goal(\neg sensation)$. Alternatively, a lack of sensation $S = Belief(\neg sensation)$ can be paired with a positive desire to experience it $D = Goal(sensation)$. This model can account for both *avoidance behavior* (e.g., polygons are destroyed when they take enough damage, and nociceptors detect damage. When the agent's arm mesh sustains a strong force, desire is signalled to stop activation of the arm's nociceptor):

$$\{S, D\} = \{Belief(pain), Goal(\neg pain)\}$$

and *approach behavior* (e.g., green-atoms are the agent's store of energy, and have a property that is detected by the mouth's sweet-taste receptors when eaten. When the agent's green-atom stores drop below critical levels, desire is signalled to activate the mouth's sweet-taste receptors):

$$\{S, D\} = \{Belief(\neg sweet), Goal(sweet)\}$$

In the avoidance case, the agent is motivated to stop an event (the pain, such as by withdrawing its arm from the painful stimulus), whereas in the approach case, the agent is motivated to realize an event (tasting sweet, such as by seeking and eating green-atom). A sensation S alone is neutral information, making the agent aware of the current situation. However, by additionally signalling motivation D, a homeostatic system influences the agent's reaction to S.

3.5 Evolving Intelligence (Implications of AIKR)

We have conjectured that a naturalistic simulation should produce increasingly prolific AGI species. However, it is easy to see where this strategy will fail. With a boxed-in environment and no energy constraints, agents will infinitely reproduce until the environment is overflowing with simple reproduction-optimizers. In the context of AGI, we want to evolve increasingly *intelligent* species, though prolific agents help prevent global extinction. According to the assumed working definition of intelligence, this means we wish to improve each individual agent's capacity to adapt to its environment (i.e., its ability to learn and execute skills on the object level) under AIKR, using the evolutionary parameters we have available (via genomic adaptation, on the meta level).

The condition of AIKR in the definition of intelligence is an important nuance, because it moves us slightly from the traditional viewpoint of intelligence as an agent's *ability to adapt* to an agent's *ability to adapt under uncertainty and resource limitations*. So to evolve greater intelligence requires pressures in the form of resource/energy limitations (i.e., insufficient resources) and uncertainty about E (i.e., insufficient knowledge). If the environment E is dangerously uncertain but the agent manages to survive it and reproduce, then the resultant species might not be reproduction-optimizers in the most simplistic sense, but rather species who are reproduction-optimizers *despite the odds*. Such species will necessarily exhibit intelligent behaviors that are linked inextricably to both their mind-body forms and the environment E in which they evolved.

In other words, since we have no fitness function to pressure greater intelligence explicitly, agents prove their superior capability in E independently by surviving and achieving reproduction. If the environment is too gentle, there is no reason for species to evolve better capabilities. If the environment is too harsh, everyone goes extinct. One problem is, any initially "difficult" static pressures that we design in the environment might be overcome by some specially-adapted species, which would then explode in population and plateau in complexity as the evolutionary pressure is nullified.

This means the environment needs to exhibit adaptive, even scaling, pressures on the global population. In this regard, the agents are their own ideal pressure, and they must be allowed to interact and even (implicitly) forced to compete. As one species improves their capability in E, another will be pressured to keep up or risk extinction. Macro-scale competition between species results in some species adapting, then others, in a neverending back-and-forth game of genetic improvements. The result is an "*arms races between and within species*" (Dawkins and Krebs, 1979), which is precisely what is needed to prevent intelligence from plateauing.

The uncertainty arising from an uncountable amount of complex evolving agent-environment and agent-agent interactions in naturalistic simulation would overwhelm traditional AI systems. For example, reinforcement learning agents assume states are repeatable, and so would fail in a naturalistic simulation which never exactly repeats. On the other hand, this treatment is possible with AGI systems working under AIKR as they are fundamentally equipped to deal with uncertainty.

3.6 Initial Conditions

Once the atoms and encoding for α are decided, the actual initial state of environment E must be created. This is up to creativity of the designer, with two notable constraints.

Firstly, E must be *contained* with finite (though renewable) resources. Consider an environment that is just an infinite 2D plane with agents and resources on top if it. Not only would this provide agents with infinite resources, but the agents might simply spread out and never reproduce. Those that do would have no reason to evolve greater abilities, as they could always migrate to a fresh new location when they use up all the simple resources in their current location. In contrast, organisms which are forced together will have no choice but to interact with each other and compete over the limited resources, thus pressuring impressive evolution via the arms race phenomenon.

Secondly, E must be populated with one or more initial organisms at $t = 0$. The organism must at minimum possess reproductive ability, motivation, and embodiment, spending energy to do work. A starting point is the most simplistic organism possible according to your specification of α; perhaps a simple body capable of movement, tactile sensors or a small eye, asexual reproductive organ, and a simplified gastrointestinal system (mouth, stomach, etc.) to maintain energy levels. Such organisms should begin to fill up the space of E, and the evolutionary takeoff begins.

4 SUMMARY

By selecting a general-purpose AI framework and applying nature-inspired principles in simulation, we can explore a huge variety of AGI species and even pressure them towards greater intelligence. Naturalistic simulations provide insights into the selected AGI model and could even yield interesting or impressive species for real-world applications.

The first step is to define the atomic building blocks of the simulation. These *atoms* provide justification in the form of *equal exchange* for structures in the environment E and the agent population. Atoms should be fairly exchanged for energy expenditure and the volume/function of *body parts* instantiated from α, such as during offspring creation. The definition of "equal" in equal exchange might be arbitrary and up to the designer, as long as the exchange rates are coherent and invariant. Just as atom classes are hardcoded, the simulated transduction of environmental signals to sensory signals seems to require hardcoded functionality, depending on which *modalities* the designer wishes to explicitly support.

The next step is to specify which aspects of the AI system and body are parameterizable (and in turn, evolvable). There are two overarching classes of parameters to evolve and encode in the genome: *cognitive parameters*, which modify the selected AGI model (e.g., its control system, architecture, decision thresholds, etc.), and *bodily parameters*, which modify the agent's embodiment (e.g., morphology, physiology [including homeostatic systems], and sensorimotor capabilities). As such, the selected AGI model must expose certain *cognitive parameters*, whereas the simulation designer specifies the available *bodily parameters*.

It is important that parameters of both types co-evolve since their interplay may be delicate when it comes to the agent exhibiting coordinated behavior. Parameterization is among the most important factors to consider when designing an evolutionary simulation, since it fundamentally constrains the range of species that can possibly be evolved. The more general and flexible these parameters are, the more opportunities for novel, interesting, and useful abilities to evolve. On the other hand, too many options in α (especially for parameters which drastically alter the core functioning of the system) can impede the search for high-performing species, as the simulation may waste computational resources on offspring with relatively incoherent mind-body parameterizations.

Motivation seeds are "externally" sourced by the agent's *body parts* during homeostatic imbalances, though such homeostatic systems will likely only be burdened by a species in exchange for increased survival and reproductive success. Sensation signals keep the agent informed of its current homeostatic outlook, while motivation signals sway the agent's behavior. At absolute minimum, reproductive motivation should be guaranteed in every organism so as to perpetuate evolution.

Despite hunger's appearance as a motivation unique to organic creatures, energy maintenance motivation (such as in the form of hunger) is likely also essential in artificial agents, since working with *finite* resources demands a work-energy tradeoff. Therefore, an agent without hunger will quickly use all its energy without replenishment and die before it can reproduce. Energy limitations force an agent to be smarter about how and when it acts so as to achieve its goals.

In many cases it is implausible to numerically measure performance on complex or abstract tasks. This is especially true when it comes to measuring an ill-defined concept like "intelligence", which has been interpreted in various ways. A naturalistic simulation is open-ended, rejecting any single choice for fitness function μ to evolve intelligent behavior. Instead, as in nature, many capable species co-evolve when individuals of each species prove their own reproductive worthiness.

Agents are pressured to evolve better adaptive capabilities when the environment contains dangerous uncertainties (increasing the importance of operating with insufficient knowledge) and their vital resources are strained (increasing the importance of working with insufficient resources). Relatively static environmental resource pressures will become less relevant as species evolve greater capabilities towards surviving in their environment. However, if the environment is contained, the species themselves should exert scaling adaptive pressures on each other as they evolve to out-compete each other for the limited resources.

Though there are many open questions, naturalistic simulation seems like a plausible way to evolve AGI agents with various intelligent capabilities. Species evolved in especially physically realistic simulations could even be instantiated in real-world robots designed analogously to the simulated bodies. A simulation can be slowed down, allowing us to observe agents interacting in real-time, or sped up, to hasten the evolutionary process and explore new species populations. Overall, naturalistic evolutionary simulations are a tool to automatically explore a wide range of capable AGI species, collect empirical data, and gain insights into both AI design and the nature of "intelligence".

REFERENCES

Canino-Koning, R., Keagy, J., and Ofria, C. (2017). Sexual selection promotes ecological speciation in digital organisms. In *ECAL 2017, the Fourteenth European Conference on Artificial Life*, pages 84–90. MIT Press.

Craig, A. D. (2003). A new view of pain as a homeostatic emotion. *Trends in neurosciences*, 26(6):303–307.

Cummings, J. A. and Becker, J. B. (2012). Quantitative assessment of female sexual motivation in the rat: Hormonal control of motivation. *Journal of neuroscience methods*, 204(2):227–233.

Dawkins, R. and Krebs, J. R. (1979). Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1161):489–511.

Fisher, H. E., Aron, A., and Brown, L. L. (2006). Romantic love: a mammalian brain system for mate choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1476):2173–2186.

Hahm, C., Xu, B., and Wang, P. (2021). Goal generation and management in nars. In *International Conference on Artificial General Intelligence*, pages 96–105. Springer.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. The University of Michigan Press.

Juliani, A., Berges, V.-P., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., et al. (2018). Unity: A general platform for intelligent agents. *arXiv preprint arXiv:1809.02627*.

Sims, K. (1994a). Evolving 3d morphology and behavior by competition. *Artificial life*, 1(4):353–372.

Sims, K. (1994b). Evolving virtual creatures. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pages 15–22.

Soros, L. and Stanley, K. (2014). Identifying necessary conditions for open-ended evolution through the artificial life world of chromaria. In *ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 793–800. MIT Press.

Strannegård, C., Engsner, N., Ferrari, P., Glimmerfors, H., Södergren, M. H., Karlsson, T., Kleve, B., and Skoglund, V. (2021). The ecosystem path to agi. In *International Conference on Artificial General Intelligence*, pages 269–278. Springer.

Strannegård, C., Xu, W., Engsner, N., and Endler, J. A. (2020). Combining evolution and learning in computational ecosystems. *Journal of Artificial General Intelligence*, 11(1):1–37.

Tsitolovsky, L. E. (2015). Endogenous generation of goals and homeostasis. In *Anticipation: Learning from the past*, pages 175–191. Springer.

von Neumann, J. (1966). *Theory of Self-Reproducing Automata*. University of Illinois Press. Edited and Completed by Arthur W. Burks.

Wang, P. (1995). *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. PhD thesis, Indiana University.

Wang, P. (2009). Embodiment: Does a laptop have a body? In Goertzel, B., Hitzler, P., and Hutter, M., editors, *Proceedings of the Second Conference on Artificial General Intelligence*, pages 174–179.

Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2):1–37.

Wang, P., Hahm, C., and Hammer, P. (2022). A model of unified perception and cognition. *Frontiers in Artificial Intelligence*, 5.

Wise, R. A. (1987). Sensorimotor modulation and the variable action pattern (vap): Toward a noncircular definition of drive and motivation. *Psychobiology*, 15(1):7–20.

Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., Lederman, S. J., and Merfeld, D. (2006). *Sensation & perception*. Sinauer Sunderland, MA.