# Pattern Survival versus Gene Survival

Randal A. Koene, *Halcyon SIM, Halcyon Molecular*
*rak.minduploading.org*
*carboncopies.org*
*www.facebook.com/randal.a.koene*

I decided to write this article after it occurred to me that many colleagues and participants whom I spoke with at the recent Humanity+ (*ref. R.A. Koene, 2010b*) and Transvision (*ref. R.A. Koene, 2010a*) conferences were struggling with personal and strategic decisions when they considered what sort of future to strive for. We are hampered by a historical dearth of attention to the very fundamentals that could support choosing a technological objective, such as cryonics, the elimination of biological aging, artificial general intelligence, or mind uploading to a whole brain emulation or other implementation of substrate-independent minds.

There is a brewing debate about whether it is truly possible to enhance the human experience, or whether the way we experience being is in fact already the most that we can aspire to. In general, we can ask: How well-considered are the different goals espoused by transhumanist thinkers? Which ones are supported by a sound rationale?

None of us want our efforts to go to waste, or to chase down lesser and near-sighted ends. Very specifically and very personally, we can ask:

> *What does a self-consistent, intelligent and capable person do? Which goals are so sound, so promising and so exciting that you can allow those goals to fully motivate you? Which goals can you embrace in the knowledge that you stand on a firm foundation, that your thinking is clear, and that you can be a pioneer to excel in a significant part of a vast new future?*

This is very important, because each of us has to choose where to dedicate our time and our effort. Similarly, solid foundations should inform decision-making about all kinds of support that can be given to specific types of projects.

In my work, I have reached this point twice, from different angles. I arrived at it once by daring to ask myself the deeper questions behind the search for greater longevity. I arrived at it the second time by questioning basic expectations proclaimed by researchers in the field now known as artificial general intelligence (AGI). I begun to address the problem from the latter angle when I spoke at the recent Winter Intelligence Conference at Oxford University (*ref. R.A. Koene, 2011*). In this article, I will therefore address the problem of solid foundations, with an emphasis on the matter of longevity... or more crucially: emphasizing the matter of survival.

## Solid context for your quest

*Well, what do you aim for?* We will need to better understand the context of the question first. Let us establish some of the bedrock rules of our universe. *There is no universal purpose.* Let go of all of the flimsy constructs that rely on notions of what should be. What we *do* observe and *can* build on is causation. One perspective that is built on causation is the concept of Universal Darwinism (*ref. D. Dennet, 2005*). We will discuss Universal Darwinism in a moment.

Above, we have the universal, objective context of the question. What is the *subjective* context? Of

course, you are not aware of the entire universal context. In fact, the only context you are aware of, contemplate and care about is the one generated by the confluence of retrieved memory, processed perception and executive processing within your own mind. That is as much of reality as there ever is to any one of us. Within that reality, that context, you choose goals, because you have interests, wishes or desires that are directly related to further possible experiences within that subjective context. Some future experiences you want to have, some you want to avoid[i].

Having established those two contexts essential to our question, I commence with a simple examination of the differences between "Gene Survival" and "Pattern Survival," their place in Universal Darwinism, and their place in our subjective interests. As I will show, the differences increasingly give us reason to drive a change in focus from the former to the latter.

## Universal Darwinism and being aware

Universal Darwinism (*ref. D. Dennet, 2005*) is a useful way to look at the results of competition throughout the universe. This extends beyond the realm of the animate, as in the biosphere of Earth. Inanimate aspects of the universe likewise experience the consequences of interactions that can be deemed competitions. When we apply this perspective, we see a tendency everywhere for some structures, some discernible components of the universe to prevail over others and, thereby to occupy a larger niche in space and time.

Likewise, it is useful to recognize that the organization of the universe, down to its quantum level, can be thought of as an arrangement that is describable, that is information (*ref. S. Lloyd, 2006*). This information universe determines all the relationships of its constituent parts, its various incarnations at different times, even the many possibilities represented by the concepts of the "multiverse" (*ref. D. Deutsch, 1997*).

When we combine both of these realizations, then we can describe the effects of Universal Darwinism in the Information Universe as a competition for "Pattern Survival." A pattern is some specific packet of information, which when put to use will achieve certain interactions and consequences.

There is a pattern that is very dear to us. This pattern is the information content of our minds. By the information content, I mean both the parameter settings (e.g., memory), as well as the ways in which the parameters are used, the functions carried out by the mind (e.g., learned behavior, characteristics) (*ref. C. Eliasmith & C.H. Anderson, 2003*). Why is this pattern so very dear to us? Well, that is based on the subjective context we identified earlier. *That pattern is all that we are aware of being.*

Self, conscious existence, is a matter of mental processes. They are the combination of perceptual processing, recall and use of memories and learning, and decision making that is affected by the mind functions that were instantiated and shaped in accordance with intrinsic drives. All of what we know, sense and experience takes place within our minds. It is these patterns that define our awareness. Those patterns are, of course, themselves the result of ongoing competition between patterns within the mind, patterns that are established, reshaped, outgrown, etc. And they are the result of evolutionary pressures that led to the development of the hardware that runs the mind. The intrinsic drives are intimately connected with those evolutionary pressures, with the survival of the genes that describe a human being.

## Having long-term interests and surviving to see them through

Our experience, therefore, leads us to place great value on the patterns that are our minds, and on the survival of those patterns, both personally and in terms of the memes we support. Our identities seek Pattern Survival. We also recognize the connection through our intrinsic drives with the "Gene Survival" that played such an important role in our native environment, the biosphere of Earth.

There are significant differences between the pursuit of gene survival and the pursuit of pattern survival. Here is an example of how these differences affect personal decisions and actions in practical terms. Do you consider yourself a hard-nosed realist? A person of practical values, of business, someone who dedicates the majority of their time to the widely accepted ideals and goals of personal and business accomplishment? Do you specialize in attaining success among your peers in terms of wealth and status? Those qualities make sense as part of a strategy with the ultimate objective of improving the odds that your **children** --- or the children in another genetic line that you are a guardian of --- will be able to **procreate** in the future. A focus on social and business success, aimed at wealth and status, but *without transhumanist objectives*, is a sensible and *self-consistent* strategy for gene survival --- even though you and your pattern of personal characteristics will terminate at your death regardless of wealth or status.

Are you, on the other hand, more concerned with the ideas, the memes, that you continually champion through your very behavior, your characteristic responses and interactions? Perhaps you do not plan to have children, and you are not primarily in charge of guaranteeing the procreation of another genetic line? Is your main interest instead drawn to the pattern of developments that you would like to see in the future, what you would consider **improvements beyond the species' status quo**? If you are a transhumanist, it is sensible to seek a strategy optimized for such pattern survival and competition. If that is your chosen objective, then, rationally, such a strategy must include work towards the transhumanist goal that can enable your pattern survival; otherwise, it is not self-consistent. Seeking strategies optimized for pattern survival of mind functions is, not coincidentally, the very definition of the objective to achieve substrate-independent minds (SIM).

We need not ask if a transhumanist would prefer to continue to exist as the same pattern or be greater than it; it is simply a fact that patterns will compete and those that best modify, adapt, and expand the domain influenced by their characteristic interactions win. To be clear, I am not talking about static pattern survival, but **pattern competition**. Pattern competition favors the personal characteristics of some, and their characteristic interactions support memes that influence future developments. A simple example: There are certain ways in which **you** would like to see the future be different from the present, which is probably distinct in some ways from how anyone else would like to see it.

## A little knowledge is a dangerous thing, but a little exploration goes a long way

*But how can we understand the original or re-implemented mind sufficiently to enable it to grow? How can you cautiously escape the human "catchment" area --- the precarious balance, where, to attain greater mental capabilities, we reach insights that remove hard-wired delusions and thereby modify our finely-tuned intrinsic reward mechanisms in a way that leads to behavior that is unfavorable to survival?*

The concept of a "catchment" area (*ref. S. Gildert, 2010*) has been described as the result of evolutionary optimization of human intelligence. Our intrinsic drives are geared to seek reward that is directly linked to gene survival. All of our actions, all of our decisions, even the way we interpret our experiences are subject to reward mechanisms that were selected in accordance with gene survival. The optimization can be considered a local maximum, surrounded by alternative modes of behavior that

were not selected for and may be less suitable guides for survival. If most of the alternatives bear detrimental risks then we can consider ourselves in many ways confined to this catchment area. It may be, that the catchment area is delicate, that it resembles a small island surrounded by a rocky landscape of possibilities, some of which could endanger our survival.

Reward mechanisms tuned by natural selection are beneficial within the existing set of goals and requirements in the human environment. As we acquire insight into our own reward mechanisms, perceive their limitations and gain the ability to modify them, there is the risk that we may promote behaviors that put our survival in jeopardy. One example would be the realization that we can maximize our ability to experience reward by setting simple goals and high rewards ("wireheading"), not unlike the lab-rat caught in a pleasure-loop by continually pressing a lever that delivers dopamine to its brain. Another example would be to modify the sense of reward that we experience when we receive the agreeable judgment of our peers in matters of social cohesion and moral values. It is true that accepted notions of right and wrong have undergone changes throughout human history, but an outright elimination of some of our basic, unquestioned drives could be more perilous if carried out without extraordinary precautions.

Here we turn to exploration and safeguards. Whole brain emulation (WBE) (*http://en.wikipedia.org/wiki/Mind_uploading*, *ref. R.A. Koene, 2006*; *A. Sandberg & N. Bostrom, 2008*) is a tool that gives you the ability to explore, such as when astronomers could first use telescopes to explore the universe. And by emulating all the relevant functions as implemented in the brain you minimize any initial differences and their potential hazards. WBE is a useful way --- though not the only one --- by which to move mind functions to another substrate, because it solves at least the problem of Access. You can carry out finely-tuned experiments, which is an opportunity that goes beyond what telescopes give astronomers.

For example, we may explore what happens if you run everything in the cortex twice as fast. Or we explore what happens if you plug in flawless memory. Whole brain emulation gives you all the basics of substrate-independent pattern survival for the mind: Continuation of the set of characteristic functions and parameters that determine how a person's interactions with the environment deploy and support memes --- characteristic interactions that affect the future.

This is an *experimental* approach by which to move from the set of constraints within one Darwinian survivor arrangement to a different set of constraints within another Darwinian survivor arrangement. Skill at doing this, at hacking minds and finding the shifts or hops required, will increase as we learn. From an art, it can become a science. We may even learn how to pre-compute the values, according to a Darwinian metric, that correspond to each of the steps of some development plan aimed at modifications of mind functions.

Some of this experimentation may be carried out through brain-computer interfaces, without whole brain emulation. *Even so, advancing substrate-independent minds (ASIM) is ultimately the only way to develop the means for human minds to escape out of and make significant strides beyond their catchment area.* ASIM is not just about making thinking things. It is not simply about longevity. It is not about remaining the same. ASIM specifically addresses the search for a feasible route and a fighting chance to play a role in the future of a Darwinian universe (*http://carboncopies.org*).

## Knowing the cause is half the cure

*Pattern survival in humans is currently being driven by gene-survival, even though the evolution of*

*humans is itself merely a byproduct of the competition for gene survival (ref. R. Dawkins, 1976). So how can one motivate pattern survival without gene survival? How can one separate the desire to procreate thought characteristics that support specific memes from the desire to procreate genes in humans?*

We will not debate competition and Darwinism here any more than we would debate gravity. These are given. We begin with the **end-result perspective**, considering that which will exist: Those things that compete successfully occupy more of space-time; the patterns of information representative of those successful things that excel at competing and developing have a great impact on the universe. Gene survival is a more narrow subset of competing patterns.

An individual may choose not to play the Darwinism game, either by not aiming at any type of pattern survival or through outright suicide. That individual is simply removed by natural selection from the pool of surviving patterns. It does not change the Darwinian outcome from the larger perspective. I posit that, finally, in terms of domains in space-time inhabited by developing patterns, the greatest part of those patterns that resulted from thinking entities will belong to those entities that transcended their equivalent of highly localized gene-survival.

We are not debating good, bad, morality, or purpose. We assume only Darwinian outcomes and try to understand the properties of those evolving, thinking entities that dominate the future. There is no universal purpose by which it would be deemed intrinsically better or worse to play this Darwinian game or to opt out. That choice already depends on **your personal characteristics**, from which a corresponding degree of competitiveness and survival may follow. For the purposes of this exercise, we do not need to concern ourselves further with the opt-outs, and instead consider the routes that belong to those likely to predominate.

It is true that from a purely practical standpoint, at present, pattern survival and gene survival are linked. But there is a shift in balance that will shortly unlink them.

## Compilers and emulators incorporate the knowledge of material things

*Can a perceiving entity that is not based on the self-replicating properties of genetic material survive over a long period of time?*

When it comes to Universal Darwinism and adaptation, it is a specific set of information, a specific piece of knowledge that is adapted to a certain niche. Adapted knowledge tends to survive within its niche in some embodiment, i.e., in some "substrate." Every time a replicator replicates, it uses non-replicating physical material to build another copy. And non-replicating knowledge can be embodied in different physical substrates each time. This way, even better survival may be achieved by consistently moving to safer substrates. The material is not crucial. Life is about knowledge. Intelligence --- whatever it means --- is about knowledge or its use and survival. It is also about an interaction with the environment. A well-adapted entity's knowledge causes its niche to maintain that knowledge or pattern.

Self-replicating properties of genetic material can be arranged in the substrate that is used to compile and emulate functions based on a pattern, even if the substrate is not (human) DNA. Genetic material carries within it the ability to enact the creation of environmental conditions that favor the replication and spread of its self-same code. The body is such an environment, aiding the genetic replicators.

Substrate-independent existence implies that one can devise compilers and emulators in various

available resources to operate using the relevant patterns in a manner that includes properties of replication, propagation, and adaptation.

We can appreciate that similar patterns may appear embodied in waves in water, in electromagnetic radiation, etc. A computer virus exists in a different substrate from ours and carries out some of the replicator functions, though it is rather parasitic and makes a home for itself in resources largely arranged for its use by others. SIM seeks not only how to extract and store patterns, but also how to engineer these flexibly implementable compilers and emulators.

Gene survival is easily annihilated due to its extreme dependence on the local environment (*ref. N. Bostrom & M.M. Cirkovic, 2008*). *Gene replication by itself will not survive for significant portions of universal time.* The major thinking survivors of the space-time envelope are the descendants of thinking entities from which substrate-independent forms emerge. Competition will emerge at some point in which the successful party will be the one that has a focus on pattern survival, and that most successfully imprints its developing patterns of thought and interaction on the future. We can be Darwinian survivors if we are adaptable and up to the admittedly great challenge of moving beyond the current limitations to our thought in terms of **access**, **interpretability**, and **capability**.

Humans have been moving towards an interest in pattern survival ever since they began to think about thinking, and since they began to explore the experience of self-awareness. We see the early consequences of this shift in the remembrance of those who have contributed memes in science, art, and the history of our species. The shift is accentuated today by organized efforts aiming specifically to accomplish the necessary transition.

## Beyond an indefensible status-quo, our rational expectations and true interests beckon

*What if a human SIM contains no information about genes, the prerequisites for survival of the pattern of the brain? The program we are currently running was evolved to and is dedicated to effective gene survival and propagation. Memes are just another tool to ensure that. Gene survival seems the very foundation of everything we are and drives us to do everything we do. What if we cannot separate from gene survival without a change dangerous to the SIM's motivation for survival? What if there is no smooth way to make the cut and escape catchment?*

An unsubstantiated worry about not being able to change with adequate caution and tentativeness is not sufficient to argue against the possibility. Until there is further cause to give substance to the specifics of these separation concerns, they express something like the uneasiness that the gods of "purpose" might strike back if we dare to change the focus in terms of which thing is being perpetuated: genes vs. minds. Therein lies the specter of the old "don't tamper with nature" argument --- and yet all progress is a function of doing exactly that. Uneasiness about re-purposing that which has emerged from gene survival (namely, our minds, our perception, our sense of personal identity and self-awareness) is not in itself a practical argument against the possibility of re-instantiating a human mind on different hardware --- and then to be able to make gradual changes.

To run the first SIM, and experimentally escape catchment, it may be necessary to glean information from DNA, body simulation, or more. Matters of *scope* and *resolution* remain to be solved for mind uploading, whole brain emulation, and substrate-independent minds. It is evident, though, that there is a severely finite combination of resolution and scope that is relevant to the human experience. Consequently, that experience can be emulated as a first step toward gradual change. As the possibility exists in principle and in practice, we must determine what the minimal scope and resolution

requirements are for the most feasible technique. To understand just how finite the scope and resolution requirements probably are, simply consider the brain as a black box with processes that relate I/O data (chemical, electromagnetic, etc.) that are not drowned in noise. It quickly becomes apparent that while the amount and rate of discernible I/O is significant by today's computing standards, it is not frightfully large.

There is no reason why we should defend the survival of characteristic genes as if they had greater purpose than the survival of our characteristic thinking. Remember: *There is no universal purpose*. There is no reason to be more attached to the sequence of nucleotides that defines the human form than to the one that defines the form of an ant. The part that is interesting to us is the emergent world of thought and perception.

## The winds of change

Are there signs of a changing emphasis in humans from gene survival to pattern survival? There is reason to believe so.

There are competitive, Darwinian pressures among thinking entities. A shift from gene survival to pattern survival is a necessary preparation for the competition between our own emergent intelligence and intelligence of another origin. That other origin could be machine intelligence without the same set of intrinsic drives, or intelligence emergent in thinking entities elsewhere in the cosmos. Greater capabilities in this competition are based on a greater understanding of one's own thinking processes, and the ability to make adaptations therein. At some point, this will demand that we move beyond the captivity within boundaries of our specific drives optimized for gene survival, our primordial reward functions. That escape can be sought through a careful transition during which competitive motivation is sustained.

Look again from the perspective of the end result: Universal Darwinism applied to thinking entities; whatever adapts and survives well. Whatever you end up creating should suit that selection. A kind of SIM will fare better in many more domains than our good old flesh and bone. In the long run, *we escape doom only by seeking to escape the catchment*.

To tackle this, do not look from past to future and think that genetic survival is the current drive and therefore we can have no route to another form that can survive. Rather, think of the future first. With reason as a guide, deduce the overall qualities of the predominant outcome. Look at what *would* thrive, and turn us into that. The next successful step will also have something that drives its survival for some period of time, even if it is not Homo sapiens' DNA. Gradual and tentative changes are the safest way to move there from what we are now, if we do not know a better approach.

In other words, advance substrate-independent minds. Start with what we have: The human brain (body too, if you like), and work from there.

Of all the transhumanist strategies, ASIM is both imbued with its originating human interests and also it most directly embraces and plays the game of competitive natural selection. We aim to base its objectives on properties that can be reasonably supposed to be those of successful competing patterns from the point of view of the end result.

*20110110, rev. 20110206*

**Acknowledgements**
*Insightful review of the early drafts of this article, as well as deeper insight into matters of possible limitations on AGI and human enhancement were graciously provided by Dr. Suzanne Gildert. Additional review was provided by Michael Andregg.*

**References**

Bostrom, N. & Cirkovic. M.M. (2008), *Global Catastrophic Risks*, Oxford University Press, Oxford, U.K.
Dawkins, R. (1976), *The Selfish Gene*, Oxford University Press, Oxford, U.K.
Dennett, D. (2005), *Darwin's Dangerous Idea*, Touchstone Press, New York, NY. pp. 352 to 360.
Deutsch, D. (1997), *The Fabric of Reality*, Penguin Books, New York, NY.
Eliasmith, C. & Anderson, C.H. (2003), *Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems*, MIT Press, Cambridge, MA.
Gildert, S. (2010), Pavlov's AI: What do superintelligences REALLY want? *Humanity+ @Caltech*. Pasadena, CA.
Koene, R.A. (2006), Scope and Resolution in Neural Prosthetics and Special Concerns for the Emulation of a Whole Brain, *2006 Workshop on Geoethical Nanotechnology*, Lincoln, VT.
Koene, R.A. (2010a), I am a 25 Watt bio-computer: What are the hacks that make us who we are?, *Transvision 2010 Conference*, Milan, Italy.
Koene, R.A. (2010b), The 25 Watt bio-computer: Lessons for Artificial Human Intelligence and Substrate-Independent Minds, *Humanity+ @Caltech*, Pasadena, CA.
Koene, R.A. (2011), Substrate-Independent Minds: Pattern Survival Agrees with Universal Darwinism, *Future of Humanity Institute Winter Intelligence Conference*, Oxford, UK.
Lloyd, S. (2006), *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*, Alfred A. Knopf, New York, NY.
Sandberg, A. & Bostrom, N. (2008), Whole Brain Emulation: A Roadmap, *Technical Report*, Future of Humanity Institute, Oxford University.

[i] This realization does not imply hedonism, because I make no claim that all the experiences you want to have are pleasurable or that all the experiences you want to avoid are painful.