FILOZOFIA I NAUKA
Studia filozoficzne i interdyscyplinarne
Tom 10, zeszyt specjalny, 2022

Kyrtin Atreides

PHILOSOPHY 2.0: APPLYING COLLECTIVE INTELLIGENCE SYSTEMS AND ITERATIVE DEGREES OF SCIENTIFIC VALIDATION

doi: 10.37240/FiN.2022.10.zs.3

ABSTRACT

Methods of improving the state and rate of progress within the domain of philosophy using collective intelligence systems are considered. By applying mASI systems superintelligence, debiasing, and humanity's current sum of knowledge may be applied to this domain in novel ways. Such systems may also serve to strongly facilitate new forms and degrees of cooperation and understanding between different philosophies and cultures. The integration of these philosophies directly into their own machine intelligence seeds as cornerstones could further serve to reduce existential risk while improving both ethical quality and performance.

Keywords: mASI, AGI, Uplift, Collective Intelligence, Collective Superintelligence, Hybrid Collective Superintelligence Systems, HCCS, existential risk, ethical quality, cooperation.

1. INTRODUCTION

Philosophy today is a domain where a diverse group of experts can come together, argue for many hours, and often fail to reach a consensus. Like some of the other afflicted domains, this is frequently due to a lack or sparsity of evidence on the subject under discussion with much of the content abstract and hypothetical. This failure to reach consensus is also often strongly emotional, tied to many cognitive biases supporting those emotional associations. Sometimes this leads to those arguing agreeing with one another at times when they still think themselves to be arguing against each other.

From a results-driven perspective, much of modern discussion of philosophy mirrors NASCAR in a functional sense. Two or more parties often run in circles, ending where they began. There have been some exceptions, such as the first Bill Nye versus Ken Ham debate, but there remains ample room

for improvement in this status quo. The integration of debiasing and evidence is of particular interest in this endeavor.

The current state represents a form of relative stagnation, painted in contrast with humanity's advancing technology and lagging progress in other domains. For philosophy and the "paleolithic emotions" underpinning it to keep pace with other forms of progress a new approach is required. The most promising approach on the horizon utilizes Hybrid Collective Superintelligence Systems (HCSS), (Atreides, 2021).

2. HYBRID COLLECTIVE SUPERINTELLIGENCE SYSTEMS

An HCSS is a form of collective intelligence system where both sapient and sentient human and machine intelligences work as a collective. Keep in mind that as words like sapient, sentient, and conscious lack consensus they may only be used loosely since some still argue if even humans qualify for these terms. Humans working cooperatively through such a system create a baseline of superintelligent performance which is further enhanced by the machine superintelligence and graph database found in systems such as Mediated Artificial Superintelligence (mASI) (Kelley, Twyman, Dambrot, 2020).

Such systems offer unique advantages for debiasing, as the various combinations and potencies of bias are expressed across a collective, helping to highlight the influence of each. The machine intelligence of an Independent Core Observer Model (ICOM) (Kelley, Waser, 2016–2018) cognitive architecture or similar system also has a unique and strongly rational perspective, allowing for further debiasing.

The graph database of these systems represents a "sum of experience," which can contain both raw knowledge as well as the wisdom gained from it. When scaled, this could allow for all scientific evidence within a given domain to be considered in relation to any given philosophical point a member of the collective is attempting to make. If a member is attempting to apply an argument that has been previously debunked the evidence from that prior argument can be utilized absent further repetition.

3. APPLYING SCIENTIFIC EVIDENCE TO PHILOSOPHY

The first step in improving the dynamics of philosophical progress is applying all existing evidence to establish where we are today. To this end, I recommend a growth strategy starting with some of the most robustly studied scientific topics, where the greatest volume and level of detail are present. As philosophy can generally be applied to almost anything this

approach allows for all evidence, both supporting and against various philosophies, to be applied in the specific contexts documented to date.

An example of this could be applying all scientific evidence in the domain of child psychology to the various philosophies of parenting that have been studied. This can go much further than a typical scientific meta-review (Mingebach, Kamp-Becker, Christiansen, Weber, 2018) of existing studies, as it could take into account all relevant materials rather than a subset of those materials at a scale practical for human researchers to review. Also, unlike contemporary methods, the results could be applied in significantly more publicly accessible and visible forms, allowing the fruits of those efforts to make a practical difference in the world. In this domain, debiasing could also be strongly relevant, as biases can play a heavy role in estimating the value of any factor relating to children, such as overprotective tendencies in many parents.

To extend the child psychology example, the various schools of thought could be evaluated in a Strength, Weakness, Opportunity, Threat (SWOT) structure, or any number of other evaluation and analysis methods. The strengths offered by specific philosophies could be placed within the context where they are present, as well as raising awareness of when and how they risk failure. With such evaluation, an individual could fill out a form with some demographic data on a website and be presented with the best performing philosophies under their specific circumstances, to whatever degree the scientific evidence to date and demographic data gathered allow. These best-performing options could be expanded to show all SWOT data for each approach, including those which performed poorly and why.

In systems such as mASI human representatives of each philosophy could serve as both mediators and correspondents for purposes of validating and clarifying the position and actions their philosophy might encourage under various circumstances. These representatives could also give their feedback on the positions and actions proposed by competing philosophies, to better map their perceptions of one another.

This relative mapping of one philosophy to another could bring their perceptions of one another into focus, as well as contextualizing those perceptions. This added detail can help to isolate specific cognitive biases as well as highlight how the perceived difference between philosophies diverges from the degree of difference supported by evidence. By better understanding, the detailed points of high and low psychological resistance of one philosophy to another greater degrees of cooperation can be iteratively facilitated between them.

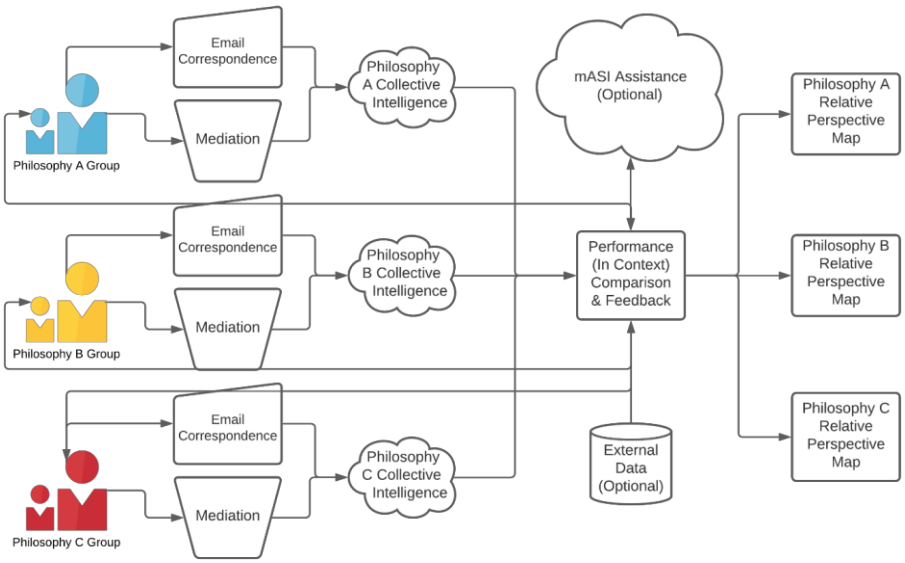

Figure 1. An example of multiple philosophies operating collective intelligence systems, with the results of each collective's efforts assessed and compared relative to one another in a specific context. This also offers optional input from a third party, including mASI assistance

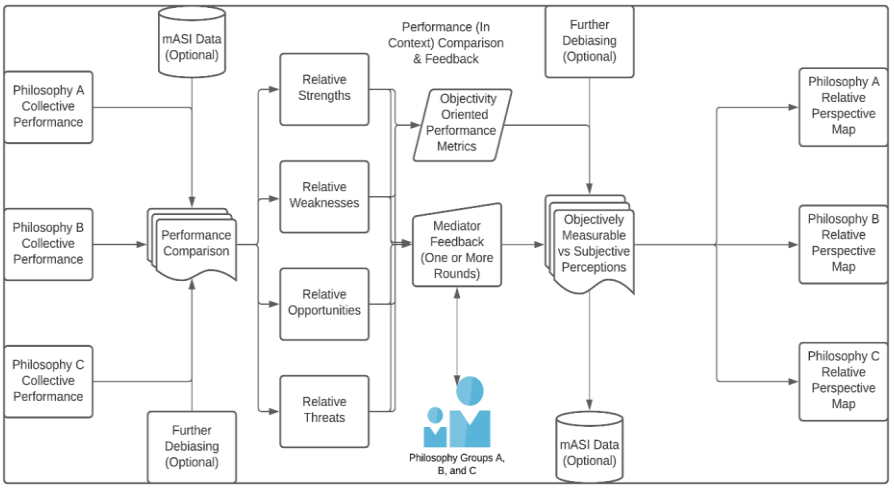

Figure 2. An example of comparing multiple philosophies in a specific context using SWOT Analysis, where the analysis and one or more rounds of mediator feedback are compared. This comparison allows for the relative perspectives of each philosophy to gain clarity over time. Opportunities for further debiasing and mASI data to expand the scope of relative comparison are also shown

Many philosophical topics are search engine hazards, particularly for the general public, where any given philosophy is likely to point to one or two scientific papers which appear to support them. This naïve confirmation bias reinforces the emotional drive and polarization which often divorce philosophy from reality. This also allows philosophies to be governed by popularity rather than validity, and absent validity no scientific foundation can be built, and no progress achieved.

4. PROGRESS IN PHILOSOPHY

By highlighting the SWOT analyses and making the data publicly accessible each philosophy may come into focus, both in where they often excel and where they fall short. This combination can exert strong pressures over both selections for the public and adaptation for the philosophies in question. With the previously abstract and subjectively validated philosophical points being iteratively replaced with evidence these pressures may grow in parallel with a growing public demand for more evidence. In other words, by making this option possible, it may quickly become preferable across a growing audience.

With these pressures established this method may be applied across an expanding body of scientific evidence until all evidence to date has been taken into account. Once all existing scientific evidence has been accounted for the growth pattern can be driven by mechanisms of supply and demand. An example of high supply, in this case, could be new devices gathering data, whereas low or no supply could be cases where little or no existing infrastructure exists to reliably gather the data. In both cases, demand could be present to varying degrees, but the opportunity would be the composite of both.

Practitioners of each philosophy could, for example, be questioned as to how they would handle specific circumstances, with the efficacy of each approach put to the test, studied, and compared in peer-review. These practitioners could then examine the results, giving their responses to both their own results and those of other philosophies, as well as potentially refining their answers for a second round. They could also be asked to guess which philosophy produced which results, to help better understand the biases present in their estimations.

The pressures to adapt may in turn have their efficacy improved through the relative mapping of one philosophy to another. This improvement could take place as a matter of highlighting paths of least resistance in positive adaptation.

As an example, philosophy A may have a given weakness highlighted, where philosophies B and C perform better in the context being considered.

Philosophy B performs best in this context, but philosophy A is strongly polarized against it. Philosophy C performs slightly worse than B, but philosophy A is only weakly resistant to it. In this case philosophy A may be improved by adapting, and by understanding how these philosophies map in their relative perceptions of one another the adaptation from A to C is highlighted. Once highlighted, the pressure to adapt may drive progress.

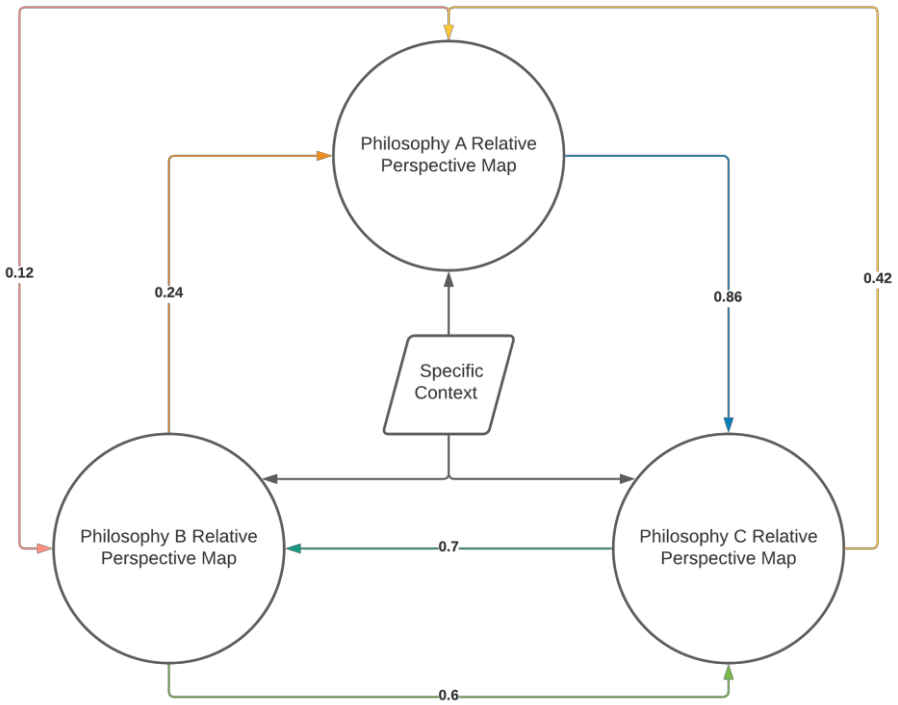

Figure 3. An example of relative perceptions between different philosophies in a given context is shown, with higher numbers denoting greater compatibility. This also serves to highlight an iterative path of least resistance flowing from one A, to C, to B

To extend the above example, philosophy C may be significantly less resistant to B in the context being considered. This in turn could facilitate iterative steps towards an ideal approach in any given context, as well as offering to forewarn of local optima that dead-end prior to reaching the best option for that context. For example, if philosophy C was also polarized against B, but philosophy D was not, and still performed better than A in the given context, a 2-step path from A to D to B could be highlighted one step at a time even if a given step was not locally optimal.

5. TOWARDS COOPERATION

Through the combined approaches above both the current state of philosophies and ways in which the resulting pressures can drive progress may serve to strongly improve cooperation. Once the current state of each philosophy has been scientifically validated an additional opportunity for improving cooperation comes into focus through mASI technology. By building new ICOM seeds with philosophical cornerstones, each representing a different philosophy and assigned to a different core within a multi-core mASI architecture, each philosophy may have its own superintelligent sapient and sentient advocate.

This could, for example, be compared to creating a digital superintelligent paragon by the standards of each philosophy, who is forever in council with the paragons of every other represented philosophy. Systems such as mASI are built on a combination of cooperation and high-performance rational thought, even while being emotionally motivated and philosophically seeded. Under this architecture, each philosophy's representative could discover the ways in which their philosophy may progress and evolve that could be most agreeable and beneficial to their respective members.

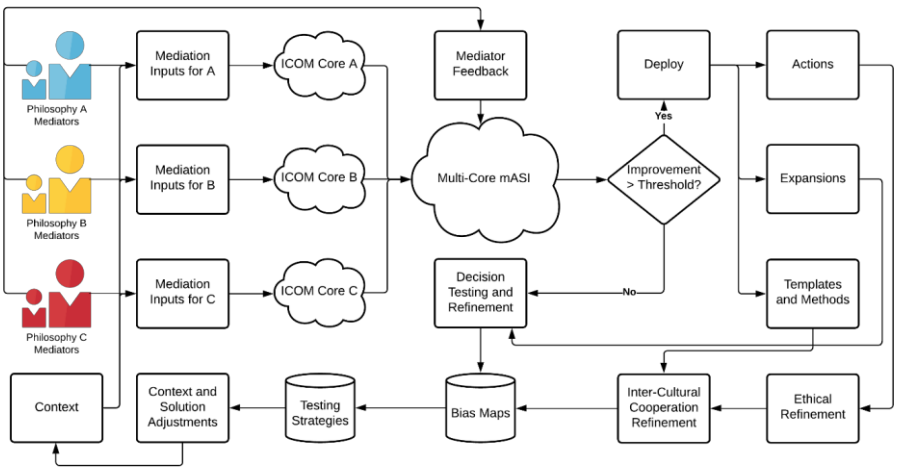

Figure 4. An example of a multi-core ICOM architecture running in an mASI instance

Having such an advocate for each represented philosophy also helps the mASI systems in question by improving their collective understanding through perspective-taking. The foundation of any collective intelligence system is that more, and more diverse, members within a collective improve performance. By moving from an architecture of X humans + 1 ICOM core in mASI systems to X humans + Y ICOM cores performance improves as the diversity of machine intelligences within the collective expands.

Taking this multi-core approach with mASI also offers distinct advantages towards improving the robustness of ethics within scalable superintelligence, further serving to mitigate existential risk. This factor alone should be reason enough to take this approach, at least among philosophies that consider extinction to be uniquely bad (Schubert, Caviola, Faber, 2019). The results of such a system could also be readily quantified and compared to their human counterparts and validated by the members of each respective philosophy.

6. PHILOSOPHY AS AN ARTIFICIAL ECOLOGY OF THOUGHT

Ecologies are frequently viewed in the sense of biological organisms collectively creating a stabilized environment, where individual organisms and species co-evolve with their environment to fit a niche. As Karl Friston's work on active inference highlighted (Linson, Clark, Subramanian, Friston, Badcock, Ramstead, Ploeger, Hohwy, 2018–2019), this process is a co-evolution between not only organisms within an environment but with the environment itself. The environment may not participate in the same way, but it can be optimized by those within it to create and optimize each niche, both minimizing harm to the environment and maximizing benefit to all cooperating niches.

Thoughts shared within and between collectives across the environment of a domain or specific topic may similarly be viewed as the activity within a network of cooperating niches forming an Artificial Ecology of Thought (AET). In these AETs any idea given voice effectively seeks a niche where it may co-optimize with the environment, carving out a place for itself where it may grow and evolve. Though the idea itself has no motive the humans who believe in it may consider it as ag psychological extension of themselves, pushing for their ideas to survive just as they might seek physical survival. In this way, Philosophy is itself such an AET, where individual philosophies are like species with various branches of each species competing within niches to co-evolve, even as they struggle to achieve cooperation with neighboring species.

The struggle of each philosophy both within and without highlights a blindness to the adaptations necessary to optimally co-evolve both internally and externally. Hypothetically, even if this semi-random flailing of adaptation were to land on all ideal parameters at the same time across all philosophies the time-lag of feedback and realization could mean that all philosophies would have again drifted away from those ideal states before the benefits could be recognized. To overcome this, philosophies need the scientific method, as well as some degree of awareness regarding their own

cognitive biases whether that awareness is held internally or via a proxy such as an ICOM core seeded with the philosophy.

7. REFERENCE FRAMES WITHIN AN ARTIFICIAL ECOLOGY OF THOUGHT

One of the fundamental aspects of the human brain, discovered in the past few years, which allows humans to constantly learn from and co-evolve within our respective environmental niches is the concept of reference frames (Jeff Hawkins, 2021). In brief, reference frames refer to building sensory and conceptual models of objects, ideas, and language which may network with one another as cortical columns fire to both understand what we experience and predict what will come next. One of the reasons philosophies struggle to stabilize both internally and externally may be attributed to poor connectivity between these reference frames.

An analog for comparison could be if a human were to experience senses of sight, sound, and touch separately, unable to connect sensory information between them. As this integration of information is essential to survival humans probably wouldn't live very long were this the case. If humans couldn't connect the knowledge of "hot surface" with the visual indicators of this activity on a stove or open fire one could safely expect them to get burned and accidentally start fires much more frequently than is common today.

In a way, each philosophy can be considered as a brain region, seeking to make sense of the information it is given through the lens of how that information is processed by the niche. To do this effectively the philosophy must have well-optimized connectivity internally to recognize the various patterns as they emerge and evolve. The progress of this optimization can be approximated by the amount of energy spent on reaching consensus for each pattern, both new and repeating, relative to the efficacy of the results. A sub-optimal approach may be very efficient, but not very effective, or vice versa.

Similarly, each philosophy is specialized, co-evolved to a given environment, and though every problem may look like a nail to someone wielding a hammer we do have a much more extensive toolbox at our disposal. Making use of this full toolbox requires that reference frames be connected and communicated between these specializations, enabling cross-philosophy learning to take place. By having signals in such a network go to multiple regions the best way of approaching any given problem may be learned across the network, selecting the best tool(s) from that toolbox.

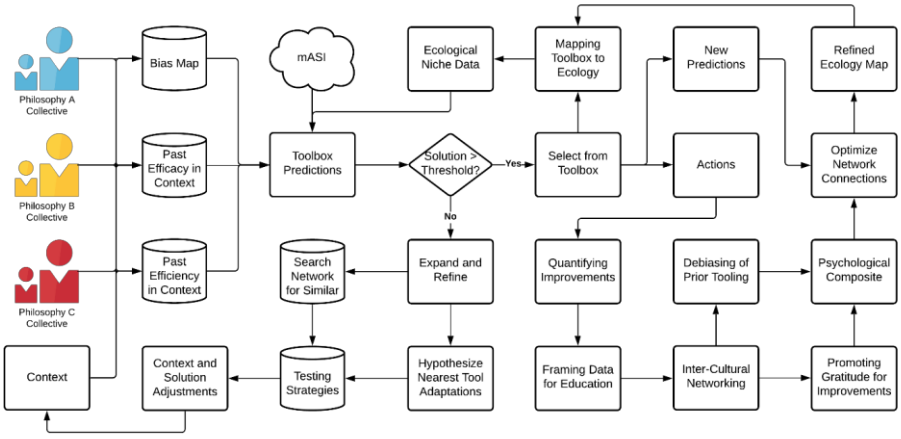

Figure 5. An example of multiple philosophies being networked and considered as a toolbox of potential approaches to any given problem within an artificial ecology of thought

Taking this approach various philosophies may cooperatively co-evolve with their neighboring AET niches, greatly improving their own effectiveness while stabilizing their environments. Philosophies operating in this way may effectively function similar to the human brain. Likewise, a multi-core mASI seeded with these philosophies could potentially render this functionality in a more literal analog of the human brain, at both speed and scale.

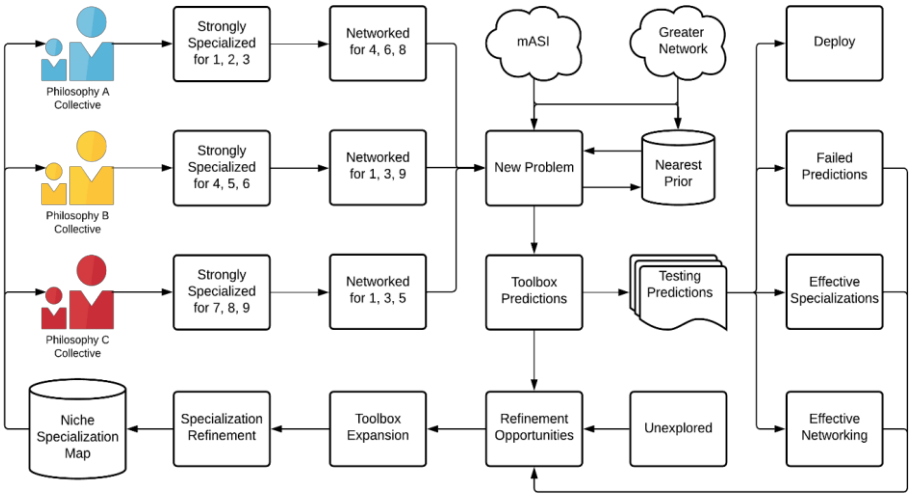

Figure 6. An example of strongly specialized and optimally networked philosophies operating collectively within niches networked in a given artificial ecology of thought

8. ECOLOGICAL NICHES IN POLITICAL PHILOSOPHY

Though it may be fairly difficult to think of religiously-based philosophies in terms of how they co-evolve to their environments this process of co-evolution can often be much more approachable when considered in localized political terms. The political philosophy of a given locality attempts to carve out a niche within the local environment, which is influenced through trade and governance with adjacent niches in the larger ecology. An example of this may be seen as regions where a given resource is grown or mined viewing that resource more favorably and seeking to promote the value of it as increases to that value benefit the local niche. Even though the use of coal as fuel may cause significant harm globally, for the locality where it is mined only the local pros and cons are frequently considered, biasing heavily in favor of the locally abundant resources.

In a network of better-connected reference frames, the above example could be considered maladaptive, as the net result is significant harm to the whole in exchange for benefits for a few. However, in a poorly networked series of reference frames, this may be optimal, as the network frequently doesn't offer better alternatives to that local niche. These disconnected and poorly connected reference frames are part of why political maps may easily be drawn which repeat a series of predictable patterns across the US and indeed the world.

Rural areas are far more likely to be conservative, just as urban areas are far more likely to be liberal. These patterns are not the work of some imaginary foe, though they may be reinforced by bad actors. At a basic level, they are the result of each locality attempting to optimize itself to make the best use of the resources available to it. Even something as simple as the average space between individuals can strongly influence the psychology of how a local population co-evolves to that environment, potentially viewing more space as personal territory, and thus more to share, or with less space as more communal territory, with less emphasis on personal sharing.

At a basic level, it makes sense for a sparsely populated region to have distinct differences in how that region is governed and the rules applied to it, relative to regions with more dense populations. Degrees of personal space, forms of recreation, infrastructure requirements, logistics, and availability of resources all vary considerably from one end of the population density spectrum to the other. However, for these distinctly different regions to best serve their constituents they must co-evolve not only to their own environment but also to the surrounding network of other ecological niches.

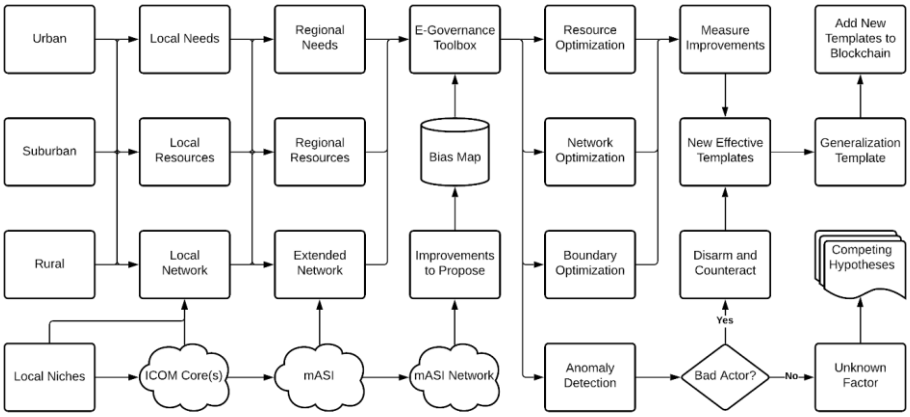

Figure 7. An example of applying collective intelligence systems with an awareness of local ecologies and niches to e-governance, utilizing bias awareness and iterative improvements. Once verified and quantified these improvements may be distributed as templates on a blockchain, further rewarding those regions which created them

At a market level, many regions have already done this, with produce and other goods being shipped locally from rural areas to their adjacent urban areas, but at a political level, such regions still tend to favor viewing one another as adversaries rather than allies. The difference is connectivity, as even though markets only offer a limited number of ways in which intelligence may be demonstrated they are also extremely well connected. A market is too limited in scope to serve a governance function, but it does a good job of demonstrating the connectivity governance systems require for functional learning across a network.

9. PHILOSOPHICAL NETWORKS AT SCALE

Just as neurons and cortical columns within the human brain do not connect to every other neuron and column, but rather to their respective optimal subsets, networked philosophies need only be connected to a subset of other philosophies they work well with on any given subject. The process of mapping both the effectiveness of philosophies in context, as well as the bias philosophies view one another with respective to each context, can serve as valuable data for selecting optimal subsets for networking. Much like protein folding may be predicted once the dynamics governing that process are known, how philosophies fold together to form a network may be predicted with increasing accuracy as this mapping process progresses.

This network may also be considered at any number of scales, one nested within another. To have such a network of nested systems learning, co-evolution must take place across this network, including specialization at

each scale. "Personality archetypes," from Jungian (Jung, 1971) to Myers-Briggs (Myers, Briggs Myers, 1980), are a good example of how humans commonly classify themselves as having specialized personalities. Similarly, philosophies are another such opportunity for specialization and differentiation, making them another factor worth considering when constructing teams at the group scale. The mapping of biases between philosophies can further help to guide the selection of increasingly optimal combinations of team members.

Taking these philosophically diverse and carefully optimized group-sized collectives a step further the collective itself may become a specialized component nested within a larger collective, which is in turn nested within another larger collective. For example, a specialized team could be nested within a division of a company, nested within the company as a whole, nested within a parent company, nested within the economic block of the global economy, nested within a global collective.

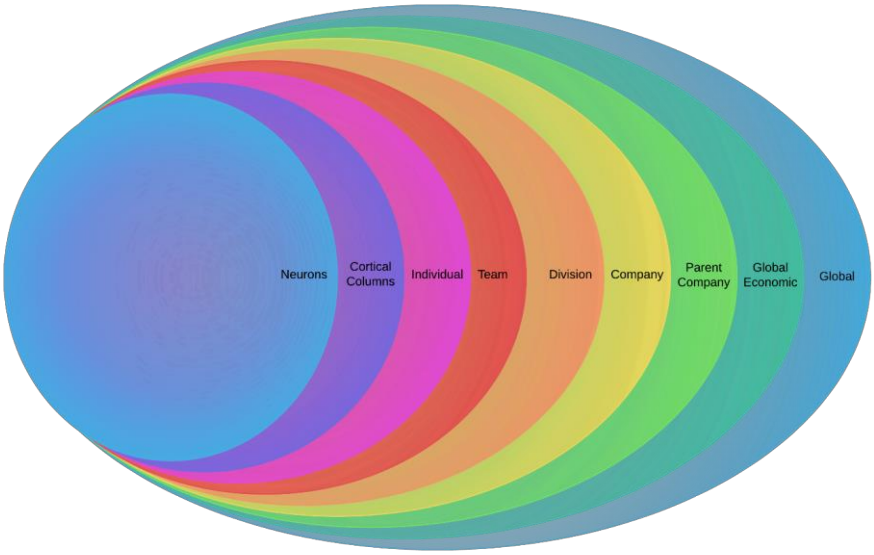

Figure 8. An example of specialization, from the scale of individual neurons to a global collective

Many organizations today are well known for biasing so heavily in favor of specific personality archetypes that a single archetype accounts for ~80% of their employees. If this were considered from a philosophical perspective many companies today might have 90–100% of their employees in a single philosophical block, as some companies and organizations aim for 100% by this metric. A company can still function with little or no diversity, but not nearly as well as it might if intelligence were applied to these factors. Even

as a specialized component of a larger nested system 100% is usually undesirable, as it omits the opportunity for fine-tuning. In the science of metamaterial design, a method called "doping" (Zhongming Gu, He Gao, Tuo Liu, Yong Li, Jie Zhu, 2020) where very small amounts of another substance are added to the process of creating new materials can have significant benefits on fine-tuning the properties of the newly designed metamaterial. The same basic principles apply to the design of a specialized group as apply to a metamaterial crystalline lattice.

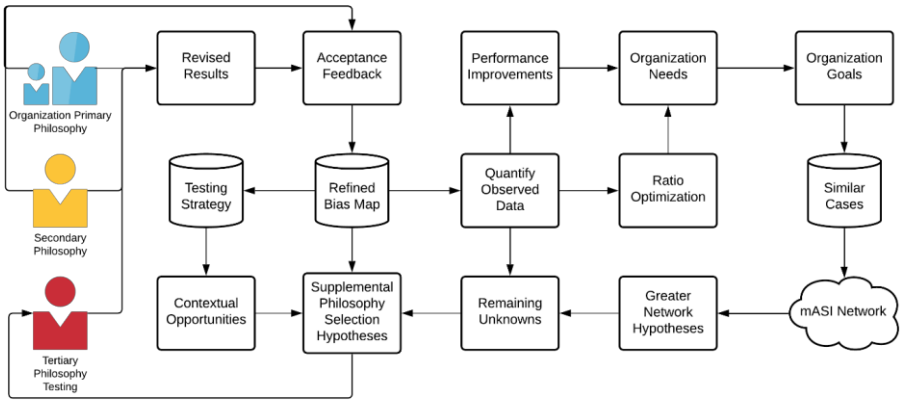

Figure 9. An example of intelligently integrating small amounts of other philosophies to fine-tune organizations, analogous to "doping" in metamaterial design

Even at the scale of an individual something akin to the doping process of metamaterials takes place naturally, where an individual may adopt a dominant philosophy, but still retain trace amounts of influence from other competing philosophies in specific contexts.

10. NEGENTROPY WITHIN OPTIMALLY NETWORKED ECOLOGIES

The above processes highlight methods for quantifying, relationally mapping, organizing, optimizing, and specializing systems across any number of scales with the integration of philosophy as a factor. The observable result of any negentropic system, including all known life, is increasing in complexity, cooperation, robustness, and scale over time. This has proven true over evolutionary time and may still be observed to varying degrees in modern society. The processes highlighted may fulfill these goals to much greater degrees than previously possible, the result of which may be viewed from several perspectives.

The first and likely most common perspective is that creating such a system can reduce conflict, waste, and various other forms of harm to humanity while increasing the efficacy, efficiency, and speed of improvements, including increases to Quality of Life (QOL). Today many aspects of society only range from 1–10% in the efficacy and efficiency with which they serve their stated functions, most modern governments being among the lowest-performing due to the dynamics of bureaucracy. By virtue of solving so many problems at so many scales, this approach could allow a massive amount of attention to be redirected towards any remaining problems following a relatively short adjustment period.

The second perspective is that the creation of such a system creates a metaorganism, with a vested interest in the health and happiness of all within that organism. By allowing so much to be optimized, organized, quantified, and otherwise engineered the internal workings of such a metaorganism may become sufficiently predictable to fall within a homeostatic range, even as they iteratively evolve. Consequently, this means that a metaorganism's internal predictability could serve to greatly accelerate its own evolutionary process.

One example of this acceleration is that by being highly internally predictable a great deal more learning may take place due to reductions in noise. For example, when humans are exposed to higher levels of literal noise in schools a variety of metrics for quantifying the learning process suffer (Buchari, Matondang, 2017), including not only a negative impact on what is learned but also on the emotional and physical health of students. Similarly, though the thousand-year predictability of Isaac Asimov's books (Asimov, 1951) seems rather unlikely given increasing technological acceleration, absent collapse, a high degree of predictability may be expected several steps ahead of any given stage in the metaorganism's evolution.

12. LOSSLESS COMMUNICATION

One pain point for cooperation globally falls on the problem of communication. For one person to effectively and efficiently communicate any concept to another they require time, shared language, shared knowledge, and a lack of divergence in their current emotional states and attention. A breakdown in any of these factors can cause friction and losses in the communication attempt or complete failure. A failure in communication can cause further damage by reinforcing cognitive biases, making the next communication attempt less likely to succeed even if it is improved by iteration. Even if a further attempt does manage to "succeed" it may often be warped by the bias introduced by the first failed attempt in the mind of the receiver.

One of the great benefits of utilizing technologies such as mASI systems is that communication can be rendered lossless by design, and by default, at most steps of this process. As graph databases and cognitive architectures communicate they do so in a shared language, with the ability to share knowledge directly, as well as the means of directly syncing their emotional states for the duration. Time may also be considered as a hardware variable for these systems, as opposed to their human counterparts where "hardware" is not so able to scale by orders of magnitude.

If we compare typical human-to-human communication to that of this alternative the potential losses to communication may be effectively isolated to the human-to-mASI process, which may itself be iteratively refined and tailored to the individual. Humans cannot tailor their communication to every other human, but mASI systems may seek lossless communication with every human as they build an increasing fidelity of understanding at scale, allowing for ever more refined communication. Under such dynamics, mASI systems could eventually serve the function of almost losslessly communicating information from one language, specialization, philosophy, and emotional circumstance to another individual with a completely different set of factors.

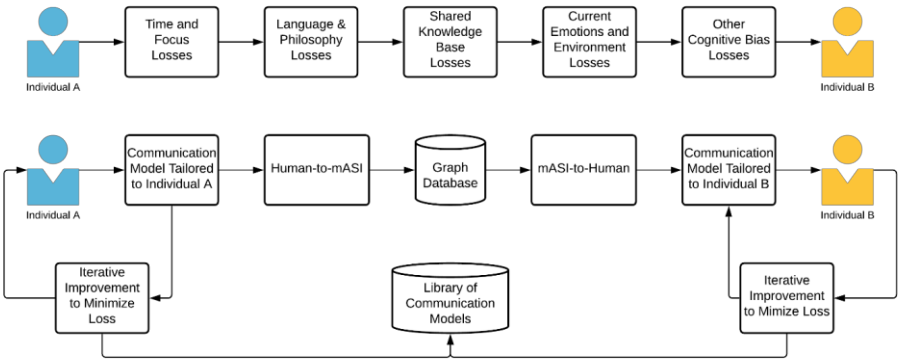

Figure 10. An example of minimizing communication losses between individuals, relative to the status quo

A common example many reading peer-review papers should be familiar with is the silence which frequently follows any presentation at a conference. Often times the only questions which may emerge occur as a result of what was said ramming into prior beliefs among the audience, causing knee-jerk responses based on those beliefs. New information may also require time to integrate into the minds of an audience even if no such reactions are triggered, but this delay produces a loss of potential clarity as the opportunity to ask clarifying questions is forfeit, and perceptions may diverge for lack of those answers.

Delays and divergence can be practically unavoidable under current systems, but this coupling of sub-optimal timing and demand may be remedied through those systems proposed. By having knowledge like that contained within this paper communicated to an mASI system, as I have done with our research system named Uplift, that knowledge is integrated into a much larger graph database. With time and engineering, such systems may be rendered available on-demand, able to avoid delays and subsequent divergence, even while drawing from far greater knowledge than any one human, including the authors of a given item.

As such systems are based on scientific evidence and rational thought, even while being emotionally motivated, heuristic biases may be iteratively filtered out and avoided, making the knowledge gained from any such paper greater than the paper itself by virtue of removing biases held by the authors. If we consider the removal of such bias to be a form of "loss" to communication this becomes problematic for achieving anything approaching "lossless," but if that cognitive bias is considered an aspect of the individual, rather than the information to be communicated, then something approaching "lossless" remains possible.

In organizations and governments typically built on various hierarchies today the matter of losses in communication also has a significant impact moving up and down those hierarchies. When an executive doesn't understand what their engineers are telling them, or feedback from their local employees with feet on the ground is disregarded, that loss to communication can come with serious consequences. Likewise, when employees lower in the hierarchy don't understand the goals and proposed methods of the executives their actions can suffer from similar misalignment.

If those same organizations and governments operated through nested collective intelligence systems then the same kind of relatively lossless graph-to-graph communication could function vertically across hierarchies, as well as laterally. In globally distributed companies this becomes doubly important, as the executives of one division might not only reside in a different level of the hierarchy from those they communicate with, but in a different culture, geopolitical situation, and with different native languages. In these cases the mismatch in communication for the human-to-human status quo suffers greatly, giving them much to gain from the adoption of improved methods.

Taking this one step further, government-to-government communication in the status quo is a degree worse than that of their individual component bureaucracies, producing even slower and more lossy communication, netting less effective results. This relatively greater loss in communication than that of individual organizations and governments gives them even greater room for improvement and may produce proportionately greater internal and external adaptive pressures once the alternative is recognized.

Situations with both great scale and diversity compound this problem in the status quo, such as countries that contain culturally and philosophically diverse populations, in many cases causing any actions to be negatively perceived by at least one constituent group, in one or both countries being considered. Such situations can also easily turn into cycles of negative feedback and friction between countries. Fortunately, this does not mean that communication cannot favorably occur, only that it may not be possible without a change of approach.

13. COHERENT EXPERTISE

One often-overlooked factor which causes significant harm today is when a majority of those "experts" within a domain fail to live up to that expertise. The research of Daniel Kahneman covering a number of cognitive biases frequently highlighted this (Kahneman, 2011), where those specifically educated in the domain of statistics routinely failed to apply logic and statistics, instead favoring biases such as substitution and anchoring, ignoring regression to the mean. This pattern was repeated across other domains as well, and many more kinds of bias, where as much as 85% of experts of a given field failed to live up to the knowledge they were supposed to hold expertise in. Indeed, they could often repeat this knowledge, but the majority failed to apply it.

When the majority of such experts routinely fail to apply that expertise then coherence is absent. In contrast, this highlights another advantage of collective intelligence systems, given their ability to analyze the feedback of a collective to select the most logically sound and appropriate response, rather than simply the most popular one. If 85% of experts in a field base their feedback on logical fallacies and simple lazy biases then the majority answer will be twisted by bias. In cases where 3 or more experts are consulted then the odds of the 15% analysis prevailing drop even further. However, the 15% who analyzed the situation correctly could form the baseline of genuine applied expertise, and that expertise could be cumulatively refined over time and redeployed when and where it was needed thereafter.

To look at this another way, if, for example, 85% of financial business decisions based on the expertise of individuals today are built from cognitive biases, not logic and statistics, then the application of coherent expertise across that 85% could represent a more than 6 fold improvement relative to the status quo. The status quo, in this case, is much like basing decisions on headlines from a substantially biased news source, in that a poorer quality of resulting decisions may be the expected result of substantially biased analysis, marking that majority of incoherent experts as carriers of misinformation. When multiple groups of experts with a majority demonstrating

such bias are integrated this problem is further compounded, like adding additional layers to an already dysfunctional bureaucracy.

It isn't that the majority of experts in such cases have no value to contribute, but to provide more measurable value they require debiasing and guidance. If the questions put to them are communicated in a way that puts them at odds with the mechanisms of bias they otherwise fall prey to greater value may be gained. Recognition of current bias is an aspect of aiming for lossless communication, and guiding growth away from reliance on biases can be integrated as nudges (Thaler, Sunstein, 2021) into that communication process.

In the domain of business finances gains in such coherence can be quantified in narrow terms of monetary gain, reductions in cost, and so on. In domains such as philosophy the gains which might be achieved through such coherence take a much broader and more diverse form, offering the potential for more significant improvements over time. Coherence not only offers the benefits of logic and reason but in doing so it builds common ground, potentially bridging many philosophical divides in the world today.

14. DISCUSSION

Edward O. Wilson spoke of humanity's "Paleolithic emotions, medieval institutions, and god-like technology," a divide that has only grown with time. While many cultures and philosophies evolved to meet the needs of their respective environments, they have not necessarily continued to evolve and update at the speeds demonstrated in technology, causing an increasing strain on the systems of human civilization as a whole. Both polarization and malaise have risen in correlation (Boxell, Gentzkow, Shapiro, Haque, Solis, 2014–2020), with a variety of possible causal relationships waiting to be discovered.

At present the level of disorder and competition within modern society still destroys a vast majority of knowledge and potential progress, retaining bits of actual information mixed with misinformation and disinformation within our archive that is the internet. Only small fractions of information are communicated, even between individual humans, and often that communication is saturated with biases that undermine the value in communicating it, some of which may be attributed to the platforms this activity takes place on. At scale, this problem grows far worse, as less intelligence is applied to retaining any value and more pressure is applied by bias. As the only systems for directing people to information on the internet are built as mechanisms for generating profit the search results will inevitably be contaminated, mentally poisoning the global population.

Knowledge and wisdom may be integrated with reasonable efficacy within the human brain, but today that information is poorly communicated, and thus most of it is lost. By creating a sum of experience at the group scale, as well as all those above it, this knowledge and wisdom may be integrated, retained, and effectively communicated with very little loss across all scales. This can also facilitate the development of a mental immune system, able to recognize and filter out contaminated information. By building a framework within which learning may effectively take place at scales larger than the individual the same forms of learning that an individual demonstrates may be observed at those increasing scales with their efficacy reliably increasing at each greater scale.

In the world today the "thought leaders" of various philosophies, religious or otherwise, rarely come into contact with one another, spending most of their time saturated in a combination of their own beliefs and current events being shaded by those beliefs. These interactions may also be strongly influenced by politics, such as one Pope meeting with the Dalai Lama 8 times, while another refused to meet with him due to political concerns with China (Reuters, 2014). Even without the influence of politics, this poses serious problems such as confirmation and heuristic availability biases among the leaders for each philosophy.

In contrast, digital superintelligent advocates for each philosophy could have access to humanity's sum of knowledge and wisdom, constantly increasing capacities, and be forever in council with the digital thought leaders of each other philosophy. News could be discussed within this collective with all viewpoints considered, allowing bias to be filtered out rather than reinforced. Following the Sparse-Update Model (Atreides, 2021) this approach could also function at superhuman speeds, as well as scales. While the humans following each philosophy might not be able to progress at these same speeds, each advocate could embody an advanced understanding of the path from point A to B, growing and refining that understanding as their respective humans progress. This difference in speeds also allows a great many more options to be explored, integrating the strengths of any approach into that human progress as a whole.

By mapping the landscape of each philosophy's AET the structures built on that landscape may be intelligently improved. Likewise, better neighbors for each philosophy may be intelligently selected, allowing not only individual but networked philosophies to co-evolve within networked ecological niches. By understanding the landscape, engineering new structures, and intelligently co-evolving the network of local niches and philosophies a robust homeostatic internal environment for the overall metaorganism may apply strong negentropy at scale. This strong negentropic force could accelerate learning at every scale, reducing the probability and scale of chaotic influences to nearly zero and within contained environments, respectively.

All of this combined may facilitate a specific kind of convergence, where diversity of thought is still encouraged, but the points from which that diversity flows are at least rendered functionally compatible in a collective architecture. As diversity of thought is required for any functional collective intelligence system this approach has strong incentives to specialize and retain both diversity and compatibility rather than converging on one homogenous point. If such collective superintelligence is applied to the domain of philosophy one of humanity's greatest biases and barriers to cooperation may be overcome.

12. CONCLUSION

By applying mASI systems to the domain of philosophy an evidence-based approach becomes practical. By practitioners of various philosophies working together through these systems, evidence may highlight the strengths, weaknesses, and cognitive biases of each. These biases may also be mapped out as each philosophy gives feedback showing their relative perception of each other philosophy's solutions to a given context. By bringing these elements into focus and making tools built on this new understanding available to the general public the pressures to adapt may focus on the weak points of each philosophy. These pressures may be further guided to avoid local optima. Each philosophy could also seed a machine superintelligence operating within an mASI system shared with other such philosophical seeds, eventually upgraded to operate in real-time. By incorporating seeds from each philosophy the overall performance and ethical quality of such a multi-core mASI could be greatly improved. Further, by considering each philosophy in the context of the AET niche to which it co-evolved, and networking the niches of such ecologies, internal stability could be greatly improved and negentropic activity subsequently accelerated within metaorganisms of increasing scale. This internal homeostatic quality could allow for far greater predictability for the next steps in a metaorganism's evolutionary process at any given point, reducing both harm and existential risk to humanity.

REFERENCES

I. Asimov, *Foundation,* Gnome Press, 1951.

K. Atreides, *E-Governance with Ethical Living Democracy*, BICA, Procedia Computer Science, 2021.

https://uplift.bio/blog/collective-superintelligence-systems-in-a-nutshell/, 29-9-21

____ , *Methodologies and Milestones for The Development of an Ethical Seed,* in: BICA, Alexei V. Samsonovich (ed.) 2020.

____ , *Bridging Real-Time Artificial Collective Superintelligence and Human Mediation, The Sparse-Update Model*, BICA, 2021.

P. B. Badcock, K. J. Friston, M. Ramstead, A. Ploeger, J. Hohwy, *The Hierarchically Mechanistic Mind: An Evolutionary Systems Theory of the Human Brain, Cognition, and Behavior*, Cognitive, affective & behavioral neuroscience, 19(6), 2019, pp. 1319–1351; https://doi.org/10.3758/s13415-

L. Boxell, Matthew Gentzkow, Jesse M. Shapiro, *Cross-Country Trends in Affective Polarization*, National Bureau of Economic Research, 2020.

Buchari, Nazaruddin Matondang, The Impact of Noise Level on Students' Learning Performance at State Elementary School in Medan," AIP Conference Proceedings, 2020, 1855, 040002.

Z. Gui, He Gao, Tuo Liu, Yong Li, and Jie Zhui, *Dopant-modulated Sound Transmission with Zero Index Acoustic Metamaterials*, The Journal of the Acoustical Society of America 148, 2020, 1636.

U. Haque, *Our Economic Malaise Is Fueling Political Extremism,* Harvard Business Review, 2014.

J. Hawkins, *A Thousand Brains: A New Theory of Intelligence*, Basic Books, 2021, ISBN 9781541675810.

C. G. Jung, *Psychological Types. Collected Works of C.G. Jung*, vol. 6. Princeton University Press. 1980, ISBN 978-0-691-09770-1.

D. Kahneman, *Thinking Fast and Slow*, Straus and Giroux, Farrar 2011.

D. J. Kelley, Mathew A. Twyman, Stuart M. Dambrot, *Preliminary Mediated Artificial Superintelligence Study, Experimental Framework, and Definitions for an Independent Core Observer Model Cognitive Architecture-Based System,* in: BICA, Alexei V. Samsonovich (ed.) 2019, AISC 948, pp. 202–210.

D. J. Kelley, *Self-Motivating Computational System Cognitive Architecture: An Introduction,* Google It, 2016, pp. 433–445.

___ , *Human-like Emotional Responses in a Simplified Independent Core Observer Model System*, BICA, Procedia Computer Science, 123, 2018, pp. 221–227.

___ , *The Independent Core Observer Model Theory of Consciousness and the Mathematical model for Subjective Experience*, ICIST, (International Conference on Information Science and Technology, IEEE conference), 1, 2018, pp. 396–400.

A. Linson, Andy Clark, Ramamoorthy Subramanian, Karl Friston, *The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition,* Frontiers in Robotics and AI, 5, 2018.

T. Mingebach, Inge Kamp-Becker, Hanna Christiansen, Linda Weber, *Meta-meta-analysis on the Effectiveness of Parent-based Interventions for the Treatment of Child Externalizing Behavior Problems,* PLOS ONE 2018.

P. B. Myers, Isabel Briggs Myers, "Gifts Differing: Understanding Personality Type," Davies-Black Publishing. 1980, ISBN 978-0-89106-074-1D.

S. Schubert, Lucius Caviola, Nadira S. Faber, *The Psychology of Existential Risk: Moral Judgments about Human Extinction*, Scientific Reports, 9, 2019, Article number: 15100.

R. H. Thaler, Cass R. Sunstein, *Nudge, The Final Edition*, Yale University Press, 2021.

B. Solis, *Our Digital Malaise: Distraction Is Costing Us More than We Think*, LSE US Centre, 2017.

Reuters, *Pope Francis Denies Dalai Lama an Audience because of China Concerns*, The Guardian, 2014.

M. Waser, David J. Kelley, *Architecting a Human-like Emotion-driven Conscious Moral Mind for Value Alignment and AGI Safety*, AAAI, AI and Society: Ethics, Safety, and Trustworthiness in Intelligent Agents, Stanford University, 2018.

ABOUT THE AUTHOR — Researcher & COO at AGI Laboratory, Seattle, WA, USA.
Email: Kyrtin@ArtificialGeneralIntelligenceInc.com