

THE EMBODIED COMMUNICATION PRIOR: A CHARACTERIZATION OF GENERAL INTELLIGENCE IN THE CONTEXT OF EMBODIED SOCIAL INTERACTION

Ben Goertzel
Novamente LLC
ben@goertzel.org

Abstract

We outline a general conceptual definition of real-world general intelligence that avoids the twin pitfalls of excessive mathematical generality, and excessive anthropomorphism.. Drawing on prior literature, a definition of general intelligence is given, which defines the latter by reference to an assumed measure of the simplicity of goals and environments. The novel contribution presented is to gauge the simplicity of an entity in terms of the ease of communicating it within a community of embodied agents (the so-called Embodied Communication Prior or ECP). Augmented by some further assumptions about the statistical structure of communicated knowledge, this choice is seen to lead to a model of intelligence in terms of distinct but interacting memory and cognitive subsystems dealing with procedural, declarative, sensory/episodic, attentional and intentional knowledge. A sister paper then extends these ideas to yield a “Cognitive Synergy Theory” that suggests specific conclusions for the architecture of artificial general intelligences, based on the ECP.

Keywords: Artificial general intelligence, human-level AI, Universal AI, learning theory, embodiment

1. INTRODUCTION

“Intelligence,” like most folk-psychology concepts, is slippery with multiple overlapping meanings, and defies straightforward formalization. Legg and Hutter [1] have reviewed over 70 published definitions of intelligence, and have proposed a fully formalized definition of their own [2]. One thing that becomes clear from their survey is that different definitions of intelligence have often been formulated with different goals in mind. Here our intent is to say something new about the topic by formulating the goal in a novel way. We aim specifically to say something about general intelligence in the sense of “embodied, intercommunicating beings who display general intelligence in the context of their interactions with each other and the world.” This is a narrower scope than many of the mathematical or philosophical definitions that have been proposed, but a broader scope than approaches that are tied specifically to intelligence that is closely humanlike.

We will begin with a very general, semi-formalized definition of intelligence that is largely in the spirit of Legg and Hutter [2], and also in the spirit of the earlier semi-formalized definition from [3]. The intuition underlying this definition is that “intelligence is the ability to achieve complex goals in complex environments.” To capture this more precisely, we may assess intelligence by averaging

achievability (by the system) over the space of all (goal, environmental situation) pairs, where the achievability of the pair is weighted by a coefficient proportional to the simplicity of the pair according to the ECP. This begs the question of how to measure simplicity – but we will defer this to the following section of the paper, as it’s one of the key original themes presented here.

So: given a system M , a goal G and a situation S , let $ach(M,G,S)$ denote the quality with which M can achieve G in S ; and let $simp(G,S)$ denote the simplicity of the pair (G,S) . Then, we may conceive the intelligence of M as the sum over (G,S) pairs of

$$ach(M,G,S) simp(G,S)$$

Fully formalizing this along the lines of [2] would be possible but would be out of place here as this is a qualitative paper that won't be pursuing rigorous proofs.

One may also define the notion of “efficient intelligence” [4] obtained by normalizing general intelligence according to computational resources utilized. There is essentially no formal theory of efficient intelligence at this point. But efficient intelligence is what matters in the real world, since theoretical general intelligences utilizing excessive amounts of computational resources can never actually be constructed.

According to this approach, different simplicity measures lead to different notions of intelligence, and hence may correspond to intelligent systems with very different properties. It seems extremely likely to us that, if one measures simplicity as program length in some standard computational model (say, a classical Turing machine; or lambda calculus), then the best approach to general intelligence is going to be some sort of simplistic brute force search algorithm such as Hutter’s AIXI [5]. Essentially, what AIXI does, at each time point, is to search the space of all computer programs to find the program P so that, if AIXI uses P to

determine its next action, it will be maximally likely to achieve its goals. After taking a single action, it then re-evaluates and chooses a new program P to execute; etc. This is a great approach if one has the computational resources to perform a thorough search of program space at each time step. Unfortunately AIXI requires infinitely much computational resources, but there is a related approach called AIXItl that restricts its search to programs with specifically bounded length and runtime, whose resource requirements are merely infeasibly humongous rather than infinite. This kind of approach has very high general intelligence, but very poor efficient intelligence. Yet, our strong suspicion is that if one measures simplicity using standard computational models, there’s probably not any way to do significantly better.

The trick to real-world general intelligence, we suggest, lies in the simplicity measure. If one assumes a simplicity measure that is biased to certain (goal, situation) pairs, then one can create intelligent systems displaying general intelligence with respect to this simplicity measure, via subtly constructing the systems in a way that embodies the patterns implicit in the simplicity measure. From this perspective, there are two interesting approaches to the “general theory of general intelligence under feasible computational resources”:

1. Discovering general principles that determine how, given a simplicity measure, to construct systems displaying general intelligence relative to that simplicity measure
2. Exploring a particular class of simplicity measures that seems to have high relevance, and figuring out how to create systems displaying general intelligence relative to simplicity measures in this class

The former would be the more ambitious approach: one can envision a theory that would let one specify a simplicity measure, and would

then produce an AI design displaying a near-maximal amount of efficient general intelligence relative to that simplicity measure. While we are interested in pursuing this direction, our focus in this paper is on the second and less ambitious goal. In the following section we will articulate a specific class of simplicity measures relevant to broadly humanlike intelligence, and then briefly explore the properties of systems displaying general intelligence relative to these simplicity measures. A sister paper [6] then explores these implications more thoroughly, developing the notion of “cognitive synergy” that emerges from some of these properties as an architectural principle for AGI design; and [7] then follows this up by looking at cognitive synergy in the specific context of the OpenCogPrime AGI architecture.

2. THE EMBODIED COMMUNICATION PRIOR

Here we describe a simplicity measure on (goal, environment) pairs; or, equivalently, a probability distribution over such pairs, which we call the Embodied Communication Prior or ECP. Here we describe the ECP intuitively rather than formalizing it rigorously; the latter would be a feasible project but would result in a much longer and more technical paper.

Consider a community of embodied agents living in a shared world, and suppose that the agents can communicate with each other via a set of mechanisms including:

- **Linguistic** communication, in a language whose semantics is largely (not necessarily wholly) interpretable based on mapping linguistic utterances into finite combinations of entities drawn from a finite vocabulary
- **Indicative** communication, in which e.g. one agent points to some part of the world or delimits some interval of time,

and another agent is able to interpret the meaning

- **Demonstrative** communication, in which an agent carries out a set of actions in the world, and the other agent is able to imitate these actions, or instruct another agent as to how to imitate these actions
- **Depictive** communication, in which an agent creates some sort of (visual, auditory, etc.) construction to show another agent, with a goal of causing the other agent to experience phenomena similar to what they would experience upon experiencing some particular entity in the shared environment
- **Intentional** communication, in which an agent explicitly communicates to another agent what its goal is in a certain situation (note interesting recent results showing that mirror neurons fire in response to some cases of intentional communication as thus defined; [8])

It is clear that ordinary everyday communication between humans possesses all these aspects. The Embodied Communication Prior is defined as the probability distribution in which the probability of an entity (e.g. a goal or environment) is proportional to the difficulty of describing that entity, for a typical member of the community in question, using a particular set of communication mechanisms including the above five modes. We will sometimes refer to the prior probability of an entity under this distribution, as its "simplicity" under the distribution.

Next, to further specialize the Embodied Communication Prior, we will assume that for each of these modes of communication, there are some aspects of the world that are much more easily communicable using that mode than the other modes. For instance, in the human everyday world:

- Abstract statements spanning large classes of situations are generally much easier to communicate linguistically
- Complex, multi-part procedures are much easier to communicate either demonstratively, or using a combination of demonstration with other modes
- Sensory or episodic data is often much easier to communicate demonstratively
- The current value of attending to some portion of the shared environment is often much easier to communicate indicatively
- Information about what goals to follow in a certain situation is often much easier to communicate intentionally, i.e. via explicitly indicating what one's own goal is

These simple observations have significant implications for the nature of the Embodied Communication Prior. For one thing they let us define multiple forms of knowledge:

- **Isolatedly declarative** knowledge is that which is much more easily communicable linguistically
- **Isolatedly procedural** knowledge is that which is much more easily communicable demonstratively
- **Isolatedly sensory** knowledge is that which is much more easily communicable depictively
- **Isolatedly attentive** knowledge is that which is much more easily communicable indicatively
- **Isolatedly intentional** knowledge is that which is much more easily communicable intentionally

This categorization of knowledge types resembles many ideas from the cognitive theory of memory [9, 10], although the distinctions drawn here are a little crisper than those which

current neurological or psychological data currently supports.

Of course there may be much knowledge, of relevance to systems seeking intelligence according to the ECP, that does not fall into any of these categories and constitutes "mixed knowledge." There are some very important specific subclasses of mixed knowledge. For instance, episodic knowledge (knowledge about specific real or hypothetical sets of events) will most easily be communicated via a combination of declarative, sensory and (in some cases) procedural communication. Scientific and mathematical knowledge are generally mixed knowledge, as is most everyday commonsense knowledge.

Some cases of mixed knowledge are reasonably well decomposable, in the sense that they decompose into knowledge items that individually fall into some specific knowledge type. For instance, an experimental chemistry procedure may be much better communicable procedurally, whereas an allied piece of knowledge from theoretical chemistry may be much better communicable declaratively; but in order to fully communicate either the experimental procedure or the abstract piece of knowledge, one may ultimately need to communicate both aspects.

Also, even when the best way to communicate something is mixed-mode, it may be possible to identify one mode that poses the most important part of the communication. An example would be a chemistry experiment that is best communicated via a practical demonstration together with a running narrative. It may be that the demonstration without the narrative would be vastly more valuable than the narrative without the demonstration. To cover such cases we may make less restrictive definitions such as

- **Interactively declarative** knowledge is that which is much more easily communicable in a manner dominated by linguistic communication

and so forth. We call these “interactive knowledge categories,” by contrast to the “isolated knowledge categories” introduced earlier.

Next we introduce an assumption we call NKC, for Naturalness of Knowledge Categories. The NKC assumption states that the knowledge in each of the above isolated and interactive communication-modality-focused categories forms a "natural category," in the sense that for each of these categories, there are many different properties shared by a large percentage of the knowledge in the category, but not by a large percentage of the knowledge in the other categories. This means that, for instance, procedural knowledge systematically (and statistically) has different characteristics than the other kinds of knowledge.

The NKC assumption seems commonsensically to hold true for human everyday knowledge, and it has fairly dramatic implications for general intelligence. Suppose we conceive general intelligence as the ability to achieve goals in the environment shared by the communicating agents underlying the Embodied Communication Prior. Then, NKC suggests that the best way to achieve general intelligence according to the Embodied Communication Prior is going to involve

- specialized methods for handling declarative, procedural, sensory and attentional knowledge (due to the naturalness of the isolated knowledge categories)
- specialized methods for handling interactions between different types of knowledge, including methods focused on the case where one type of knowledge is primary and the others are supporting (the latter due to the naturalness of the interactive knowledge categories)

3. LEVELS OF GENERAL INTELLIGENCE

Another, complementary way to explore the relationship between the ECP and real-world general intelligence, is to tie it in with various relevant theories of the *levels* of general intelligence.

Piaget [11] introduced a series of cognitive developmental stages, which may be approximately summarized as

1. **Infantile:** learns to carry out procedures oriented toward achieving goals in relevant contexts
2. **Operational:** learns abstractions allowing it to adapt its learning to different contexts and subgoals
3. **Formal:** learns abstractions allowing it to adapt its adaptation

Piaget also introduced a “pre-operational” stage between the infantile and operational ones; and some more recent thinkers have introduced a “post-formal” stage involving the ability of the system to apply formal reasoning to its own basic structure and outlook. But these three phases will suffice for our current purposes.

Taking more of a cybernetics and general systems theory approach, Gregory Bateson [12] proposed a different sort of hierarchy:

1. Learning
2. Learning how to learn
3. Learning how to learn how to learn

Of course, there is no reason the hierarchy of levels of learning needs to stop at that point; but Bateson suggests that in actual human practice it generally does.

In [13] it is suggested that these two hierarchies, which look quite different on the surface, are actually closely aligned, so that in certain types of intelligent systems, the Piagetan and Batesonian stages correspond closely.

Specifically it is argued that this correspondence holds both in humans and in AI systems whose operations are centrally based on uncertain inference. What we suggest here is that this correspondence is actually a conceptual consequence of the Embodied Communication Prior.

In the context of the ECP, the operational stage may be viewed as requiring a thorough integration of declarative and procedural learning. Declarations about procedures must be learned, and then manipulated in order to lead to new procedures. This is precisely Batesonian “learning how to learn.” Of course sensory knowledge must also be drawn into the picture here, as these declarations will often be highly dependent on sensorially-defined contexts.

On the other hand, the formal stage requires a yet deeper integration, in which procedures are learned for controlling the application of declarative knowledge to guiding procedure learning – and declarative knowledge is then learned regarding these higher-level procedures. This is a direct translation of the ideas of [13] into the language of the ECP; and clearly it is an instance of Batesonian “learning how to learn how to learn.”

Thus, we suggest that in the context of the ECP, Bateson’s third level of learning and Piaget’s formal stage come together in a dynamic of “mixed declarative/procedural learning about declarative learning about procedural learning.” And this dynamic, we propose, is critical to the achievement of a high level of efficient general intelligence according to the ECP.

4. THE COGNITIVE SCHEMATIC

Now we explore a little more deeply what the above conclusions imply about the internal operations of AI systems displaying general intelligence with respect to the ECP (and adopting the NKC assumption as well). To

simplify analysis we consider an AI system that possesses a specific set of goals, together with a predefined set of actuators and sensors. We further assume that the action of the system may be modeled in terms of the “cognitive schematic”

$$\text{Context \& Procedure} \rightarrow \text{Goal } \langle p \rangle$$

interpreted to mean “If the context C appears to hold currently, then if I enact the procedure P, I can expect to achieve the goal G with certainty p.” A procedure is defined as some systematic pattern of activity within the system (which may involve activation of the external actuators, or may in some cases involve purely internal activities). A context is defined as a fuzzy logical predicate that holds, to a certain extent, during each interval of time. A goal is simply some fuzzy logical predicate that has a certain value at each interval of time, as well. We will also use the shorthand

$$C \& P \rightarrow G \langle p \rangle$$

Note that we don’t assume the system explicitly uses the cognitive schematic to regulate its activities (nor uses fuzzy logical predicates, or the other apparatus introduced above, in its internal operations); rather, we are introducing this as an external model of the system. If the system explicitly uses some form of the cognitive schematic in its internal operations, that’s fine, but our analysis does not require this.

This formalization leads to a conceptualization of the internal action of an intelligent system as involving two “key learning processes”:

1. Estimating the probability p of a posited $C \& P \rightarrow G$ relationship
2. Filling in one or two of the variables in the cognitive schematic, given assumptions regarding the remaining variables, and directed by the goal of

maximizing the probability of the cognitive schematic

Again, we stress that we don't assume the system's internal dynamics are explicitly oriented around these two types of activity. What we assume is that the system can be modeled this way, basically as a combination of:

1. Evaluating conjectured relationships between procedures, contexts and goals ("analysis")
2. Conceiving novel possible relationships between procedures, contexts and goals ("synthesis")

Given this conceptualization, we can see that, where synthesis is concerned,

- Isolatedly procedural knowledge can be useful for choosing P, given fixed C and G
- Isolatedly sensory knowledge, isolatedly declarative knowledge, or mixed sensory/declarative knowledge can be useful for choosing C, given fixed P and G
- Isolatedly declarative knowledge can be useful for choosing G, given fixed P and C

On the other hand, where analysis is concerned:

- Isolatedly declarative knowledge can be useful for estimating the probability of the implication in the schematic equation, given fixed C, P and G. Episodic knowledge can also be useful in this regard, via enabling estimation of the probability via simple similarity matching against past experience.
- Isolatedly procedural knowledge can be useful for estimating the probability of the implication $C \& P \rightarrow G$, in cases

where the probability of $C \& P1 \rightarrow G$ is known for some P1 related to P

- Isolatedly declarative or isolatedly sensory knowledge can be useful for estimating the probability of the implication $C \& P \rightarrow G$, in cases where the probability of $C1 \& P \rightarrow G$ is known for some C1 related to C
- Isolatedly declarative knowledge can be useful for estimating the probability of the implication $C \& P \rightarrow G$, in cases where the probability of $C \& P \rightarrow G1$ is known for some G1 related to G

We can also see the role of mixed knowledge here, because sometimes the best way to handle the schematic equation will be to fix only one of the terms. For instance, if we fix G, sometimes the best approach will be to collectively learn C and P (for instance, using evolutionary learning methods, to allow them to "co-evolve"). Dominantly procedural knowledge, for example, corresponds to the case where one mainly wants to learn P, but accepts that one may also need to adapt C or G during the learning process, rather than leaving them fixed.

The final fact we need to account for is that, in any real-world context, a system will be presented with a huge number of possibly relevant analysis and synthesis problems. Choosing which ones to explore is a difficult cognitive problem in itself – a problem that also takes the form of the cognitive schematic, but where the procedures are internal rather than external. Thus this problem may be addressed via the analysis and synthesis methods describe above. This is the role of attentional knowledge; it gives the system some base knowledge regarding what to attend to (which in some cases will be the problem of using complex analysis and/or synthesis to figure out what to attend to). Attentional knowledge may be built up analytically or synthetically, and isolatedly or interactively, just like other types of knowledge. The NKC suggests that

attentional knowledge forms a natural category just like the other types of knowledge.

Suppose we conceive an AI system as consisting of a set of learning capabilities, each one characterized by three features:

- One or more **knowledge types** that it is competent to deal with, in the sense of the two key learning problems mentioned above
- At least one **learning type**: either analysis, or synthesis, or both
- At least one **interaction type**, for each (knowledge type, learning type) pair it handles: “isolated” (meaning it deals mainly with that knowledge type in isolation), or “interactive” (meaning it focuses on that knowledge type but in a way that explicitly incorporates other knowledge types into its process), or “fully mixed” (meaning that when it deals with the knowledge type in question, no particular knowledge type tends to dominate the learning process).

Then, it seems to follow from the ECP with NKC that systems with high efficient general intelligence should have the following properties, which collectively I’ll call “cognitive completeness”:

- For each (knowledge type, learning type, interaction type) triple, there should be a learning capability corresponding to that triple.
- Furthermore the capabilities corresponding to different (knowledge type, interaction type) pairs should have distinct characteristics (since according to the NKC the isolated knowledge corresponding to a knowledge type is a natural category, as is the dominant knowledge corresponding to a knowledge type)
- For each (knowledge type, learning type) pair (K,L), and each other knowledge

type K1 distinct from K, there should be a distinctive capability with interaction type “interactive” and dealing with knowledge that is interactively K but also includes aspects of K1

Furthermore, it seems intuitive that according to the ECP with NKC, if the capabilities mentioned in the above points are reasonably able, then the system possessing the capabilities will display general intelligence relative to the ECP. Thus we arrive at the hypothesis that

Under the assumption of the Embodied Communication Prior (with the Natural Knowledge Categories assumption), the property above called “cognitive completeness” is necessary and sufficient for efficient general intelligence at the Piagetan formal level.

Of course, the above considerations are very far from a rigorous mathematical proof (or even precise formulation) of this hypothesis. But we are presenting this here as a conceptual hypothesis, in order to qualitatively guide R&D and also to motivate further, more rigorous theoretical work.

An approach to AGI architecture called “Cognitive Synergy Theory” [6] goes into more detail regarding the types of cognitive process involved in intelligent systems modeled by the cognitive schematic, and the ways in which they may interact with each other and support each other. [7] then looks at the relationship of Cognitive Synergy Theory with a specific AGI architecture, OpenCogPrime.

5. BEYOND THE EMBODIED COMMUNICATION PRIOR

One interesting direction for further research would be to broaden the scope of the inquiry, in a manner suggested above: instead of just looking at the ECP, look at simplicity measures in general, and attack the question of

how a mind must be structured in order to display efficient general intelligence relative to a specified simplicity measure. This problem seems unapproachable in general, but some special cases may be more tractable.

For instance, suppose one has

- a simplicity measure that (like the ECP) is approximately decomposable into a set of fairly distinct components, plus their interactions
- an assumption similar to NKC, which states that the entities displaying simplicity according to each of the distinct components, are roughly clustered together in entity-space

Then one should be able to say that, to achieve efficient general intelligence relative to this decomposable simplicity measure, a system should have

- distinct capabilities corresponding to each of the components of the simplicity measure
- interactions between these capabilities, corresponding to the interaction terms in the simplicity measure

With copious additional work, these simple observations could potentially serve as the seed for a novel sort of theory of general intelligence – a theory of how the structure of a system depends on the structure of the simplicity measure with which it achieves efficient general intelligence. Cognitive Synergy Theory would then emerge as a special case of this more abstract theory.

6. CONCLUSION

We have introduced a novel approach to defining real-world general intelligence, attempting to occupy the middle ground between excessive mathematical generality and excessive anthropomorphism. While

lacking true mathematical rigor, the ideas presented have been semi-formalized, in an attempt to maintain a high level of conceptual precision; but nevertheless, they must be considered at this stage to overlap as much with the category of “philosophy of mind” as with mathematics or science.

Yet, we do not consider these ideas pragmatically irrelevant due to their somewhat philosophical nature. Given the current lack of a rigorous mathematical and scientific theory of real-world general intelligence, those of us concerned with constructing AGI systems or analyzing human intelligence require some sort of guidance. And, it seems better to draw guidance from conceptually clear, only partially rigorous thinking about the right issues, than from more rigorously mathematically or empirically grounded thinking about the wrong issues.

As the discussion here has hopefully made clear, theorizing about “fully general intelligence” is not likely to be useful for understanding real-world general intelligence. And yet, one would like to be able to say something about general intelligence going beyond the description of particular systems like human brains or particular software systems. On the other hand, from an AGI point of view, studying the human brain and mind is valuable, but can also be confusing and misleading, if one’s goal is not to precisely emulate human intelligence but rather to make a different, perhaps in some respects better, sort of intelligence, operating in the same everyday world as humans.

The approach taken here seeks to find a middle path, via qualitatively characterizing the class of systems that are generally intelligent with respect to the Embodied Communication Prior (with the Natural Knowledge Categories assumption); a class which, intuitively, appears to include both human brains, and the systems that would result from fully implementing and teaching certain contemporary AGI designs such as OpenCogPrime.

One promising future research direction would be to attempt to create a fully mathematically rigorous version of the ideas presented here – a quest that would doubtless involve a number of refinements and revisions of the ideas, though I suspect that the spirit would remain intact. Another, noted above, is to extend the scope of these ideas more thoroughly beyond the ECP to deal with more general classes of simplicity measures. One could also attempt to go further in the direction of directly deriving AGI designs from these theoretical considerations (rather than using the theory to analyze existing AGI designs, as done in [7]), or using these considerations to analyze human brain function. All these avenues seem valuable, along with others; clearly the study of general intelligence is still at a very early stage, theoretically as well as pragmatically.

References

- [1] Legg, S. and Marcus Hutter. (2007) A Collection of Definitions of Intelligence. In B. Goertzel, editor, *Advances in Artificial General Intelligence*, IOS Press
- [2] Legg, S. and Marcus Hutter. (2006). A Formal Measure of Machine Intelligence. In *Proc. Annual machine learning conference of Belgium and The Netherlands (Benelearn-2006)*. Ghent, 2006.
- [3] Goertzel, Ben (1993). *The Structure of Intelligence*. Springer.
- [4] Goertzel, Ben (2006). *The Hidden Pattern*. BrownWalker.
- [5] Hutter, Marcus (2005). *Universal AI*. Springer.
- [6] Goertzel, Ben (2009). *Cognitive Synergy: A Universal Principle of Feasible General Intelligence?*, *Dynamical Psychology*
- [7] Goertzel, Ben (2009a). *OpenCog Prime: A Cognitive Synergy Based Architecture for Artificial General Intelligence*. *Dynamical Psychology*.
- [8] Fogassi, Leonardo, Pier Francesco Ferrari, Benno Gesierich, Stefano Rozzi, Fabian Chersi, Giacomo Rizzolatti (2005). Parietal lobe: from action organization to intention understanding. *Science* 308: 662-667.
- [9] Eichenbaum (2002). *The Cognitive Neuroscience of Memory*. Oxford University Press.
- [10] Tulving and Craik (2005). *The Oxford Handbook of Memory*. Oxford University Press.
- [11] Piaget, Jean. (1955) *The Construction of Reality in the Child*. Routledge and Kegan Paul.
- [12] Bateson, Gregory (1980). *Mind and Nature: A Necessary Unity*. Bantam
- [13] Goertzel and Bugaj (2007). Stages of Cognitive Development in Uncertain Inference Based AI Systems. In B. Goertzel, editor, *Advances in Artificial General Intelligence*, IOS Press