COGNITIVE SYNERGY: A UNIVERSAL PRINCIPLE FOR FEASIBLE GENERAL INTELLIGENCE?

Ben Goertzel
Novamente LLC
ben@goertzel.org

Abstract

Do there exist general principles, which any system must obey in order to achieve advanced general intelligence using feasible computational resources? Here we propose one candidate: "cognitive synergy," a principle which suggests that general intelligences must contain different knowledge creation mechanisms corresponding to different sorts of memory (declarative, procedural, sensory/episodic, attentional, intentional); and that these different mechanisms must be interconnected in such a way as to aid each other in overcoming memory-type-specific combinatorial explosions.

Keywords: Artificial general intelligence, human-level AI, emergence, systems theory

1. INTRODUCTION

What can one say, in general, about general intelligence? The answer seems to be: a fair bit of interesting mathematics which has worthwhile philosophical implications, but not much of direct practical value. The tradition of Solomonoff induction [1,2], pursued recently by Juergen Schmidhuber [3], Marcus Hutter [4] and others, seems to say what there is to be said about truly general intelligence. There are many more mathematical details to be unraveled, but the conceptual picture seems

clear. Without some sort of special assumptions about the environment and goals relevant to an intelligent system, the best way to achieve general intelligence -- in the sense of generally effective goal-achieving behavior, for complex goal/environment combinations -- is essentially to carry out some form of brute-force search of the space of possible behavior-control programs, continually re-initiating the search as one's actions lead to new information. The problem with this approach is that it's incredibly computationally expensive -- which leads to the question of the best way to achieve reasonable levels of general intelligence given feasible computational resources. Here the picture is less clear mathematically at this point, but, intuitively, it seems fairly clear that achieving useful levels of truly general intelligence (without special assumptions about the world) using feasible computational resources is just not possible.

But what if one restricts the scope of generality, via making appropriate special assumptions about the goal/environment combinations of interest? In this case, one's claim for generality of intelligence is less, but if the restrictions are not too severe, one still has a case of interest. But the question is whether there is anything of elegance and scope to say about this case. Plausibly, this case might degenerate into a collection of highly specialized statements about particular classes of goals and environments. On the other hand, maybe there is something sensible

one can say about "general intelligence in the everyday world using feasible computational resources," as a topic more restrictive than mathematically general intelligence, but less restrictive than task-specific intelligence like chess-playing, car-driving, biological data analysis, etc.

Our goal here is to sketch some ideas that we think can serve as the core of a reasonably general theory of everyday-world general intelligence using feasible computational resources. Toward this end we deal specifically with the case of "multi-memory systems," which we define as intelligent systems whose combination of environment, embodiment and motivational system make it important for them to possess memories that divide into partially but not wholly distinct components corresponding to the categories of:

- Declarative memory
- Procedural memory (memory about how to do certain things)
- Sensory and episodic memory
- Attentional memory (knowledge about what to pay attention to in what contexts
- Intentional memory (knowledge about the system's own goals and subgoals)

In [5] we present a detailed argument as to how the requirement for a multi-memory underpinning for general intelligence emerges from certain underlying assumptions regarding the measurement of the simplicity of goals and environments; but the points made here do not rely on that argument. What they do rely on is the assumption that, in the intelligence in question, the different components of memory are significantly but not wholly distinct. That is, there are significant "family resemblances" between the memories of a single type, yet there are also thoroughgoing connections between memories of different types.

Cognitive Synergy Theory, if correct, applies to any AI system demonstrating intelligence in the context of embodied, social communication. However, one may also take the theory as an explicit guide for constructing AGI systems; and [6] describes one AGI architecture, OpenCogPrime, designed in such a way.

2. ESSENTIAL PRINCIPLES OF COGNITIVE SYNERGY THEORY

The essential idea of cognitive synergy theory may be expressed in terms of the following points:

1. Intelligence, relative to a certain set of environments, may be understood as the capability to achieve complex goals in these environments.
2. With respect to certain classes of goals and environments, an intelligent system requires a "multi-memory" architecture, meaning the possession of a number of specialized yet interconnected knowledge types, including: declarative, procedural, attentional, sensory, episodic and intentional (goal-related). These knowledge types may be viewed as different sorts of pattern that a system recognizes in itself and its environment.
3. Such a system must possess knowledge creation (i.e. pattern recognition / formation) mechanisms corresponding to each of these memory types. These mechanisms are also called "cognitive processes."
4. Each of these cognitive processes, to be effective, must have the capability to recognize when it lacks the information to perform effectively on its own; and in this case, to dynamically and interactively draw information from

2

knowledge creation mechanisms dealing with other types of knowledge

5. This cross-mechanism interaction must have the result of enabling the knowledge creation mechanisms to perform much more effectively in combination than they would if operated non-interactively. This is "cognitive synergy."

6. The activity of the different cognitive processes involved in an intelligent system may be modeled in terms of the schematic implication "Context & Procedure ➔ Goal", where the Context involves sensory, episodic and/or declarative knowledge; and attentional knowledge is used to regulate how much resource is given to each such schematic implication in memory

These points are implicit in the systems theory of mind given in [7] but are not articulated in this specific form there.

Interactions as mentioned in Points 4 and 5 are the real conceptual meat of CST. One way to express the key idea here is that most AI algorithms suffer from combinatorial explosions: the number of possible elements to be combined in a synthesis or analysis is just too great, and the algorithms are unable to filter through all the possibilities, given the lack of intrinsic constraint that comes along with a "general intelligence" context (as opposed to a narrow-AI problem like chess-playing, where the context is constrained and hence restricts the scope of possible combinations that needs to be considered). In an AGI architecture based on CST, the different learning mechanisms must be designed specifically to interact in such a way as to palliate each others' combinatorial explosions – so that, for instance, each learning mechanism dealing with a certain sort of knowledge, must synergize with learning mechanisms dealing with the other sorts of knowledge, in a way that decreases the severity of combinatorial explosion.

One prerequisite for cognitive synergy to work is that each learning mechanism must recognize when it is "stuck," meaning it's in a situation where it has inadequate information to make a confident judgment about what steps to take next. Then, when it does recognize that it's stuck, it may request help from other, complementary cognitive mechanisms.

Next, drilling a little deeper into Point 3 above, one arrives at a number of possible knowledge creation mechanisms (cognitive processes) corresponding to each of the key types of knowledge. Figure 1 below gives a high-level overview of the main types of cognitive process considered in the current version of Cognitive Synergy Theory, categorized according to the type of knowledge with which each process deals. Next, Tables 2 and 3 exemplify the memory types and cognitive processes from Figure 1 in the context of AI systems acting in a simple virtual world according to the "AGI Preschool" methodology described in [8]. For more thorough characterizations of these ideas, see [7].

3. THE COGNITIVE SCHEMATIC

Point 6 in the above summary of Cognitive Synergy Theory describes how the various cognitive processes involved in intelligence may be understood to work together via the "cognitive schematic"

Context & Procedure ➔ Goal <p>

This formula may 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." The system is initially supplied with a set of high-level goals such as "get rewarded by my teacher", "learn new things" and so forth; and it then uses

inference (guided by other cognitive mechanisms) to refine these initial goals into more specialized subgoals. As noted above, we use the term "intentional knowledge" to refer to the system's knowledge of its goals and subgoals. In the following will also use the shorthand

$$C \ \& \ P \ \rightarrow G \ <p>$$

Declarative Memory

Cognitive Processes Associated with Types of Memory

Sensorimotor Memory

Uncertain Inference:
deduction, induction, abduction, etc.

Unsupervised Pattern Mining

Concept creation:
Including blending

Modality specific memory :
Body map for haptics & kinesthetics, hierarchical memory for vision, etc..

Specialized pattern recognition:
Creates patterns linking modality-specific stores into declarative, procedural and episodic memory

Attentional Memory & System Control

Dynamic attention allocation:
Dynamically determining the space and time resources allocated to memory items, for resource allocation & credit assignment

Map formation
Identification and reification of global emergent memory patterns

Goal System
Refinement of given goals into subgoals; allocation of resources among goals

Procedural Memory

Supervised program learning
Learning of a program given a "fitness function"

Deliberative planning
Done in an uncertainty-savvy way

Episodic Memory

Internal Simulation
of historical and hypothetical external events
Spacetime interface:
special mechanisms for linking spatiotemporal experiential knowledge with delcarative and procedural knowlege
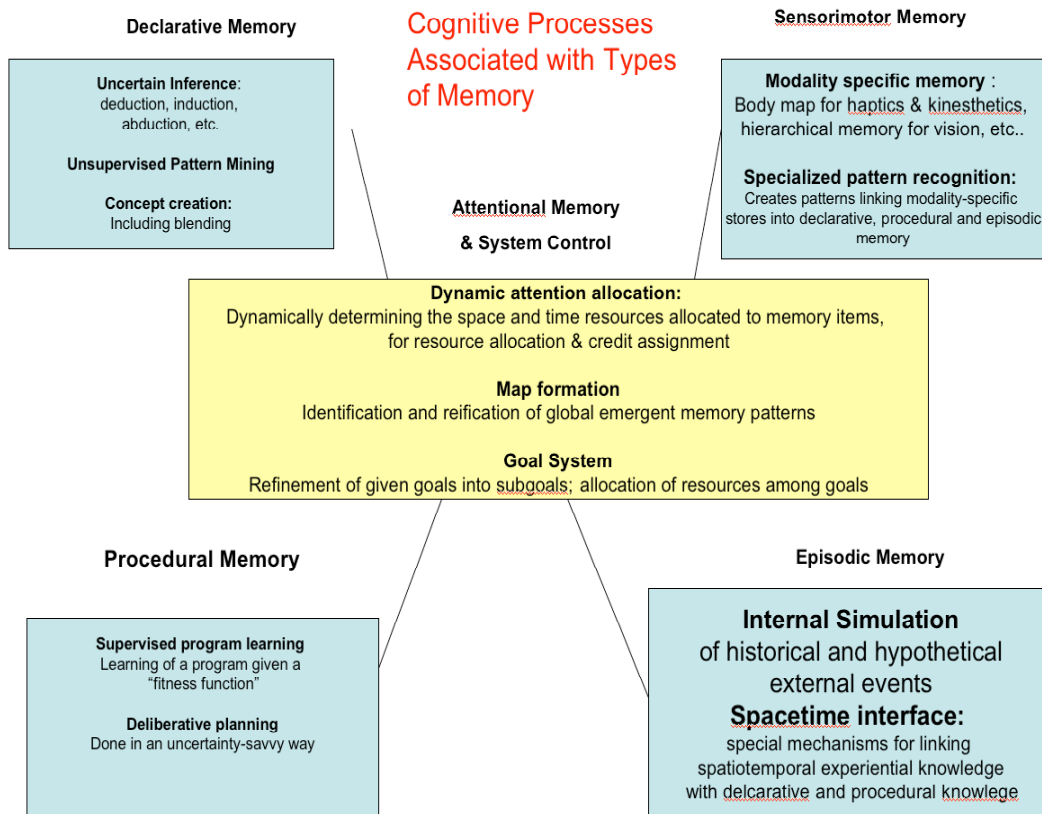
Figure 1. High-level overview of the key cognitive dynamics considered in the current version of Cognitive Synergy Theory. Cognitive Synergy Theory in its current form describes the behavior of a system as it pursues a set of goals, which are then refined by inference, aided by other processes. Terms like "inference" are used very broadly here; for instance there is no commitment to explicit use of a logic engine and, from the point of view of a high-level description like this diagram, inference could just as well be carried out as an emergent process resulting from the dynamics of an neural net system. At each time the system chooses a set of procedures to execute, based on its judgments regarding which procedures will best help it achieve its goals in the current context. These procedures may involve external actions (e.g. involving conversation, or controlling an agent in a simulated world) and/or internal cognitive actions. In order to make these judgments it must effectively manage declarative, procedural, episodic, sensory and attentional memory, each of which is associated with specific algorithms and structures as depicted in the diagram. There are also global processes spanning all the forms of memory, including the allocation of attention to different memory items and cognitive processes, and the identification and reification of system-wide activity patterns (the latter referred to as "map formation").

4

Knowledge Type	Virtual Agent Example(s)
Declarative	• The red ball on the table is larger than the blue ball on the floor • Bob becomes angry quickly • Ball roll. Blocks don't. • Jim knows Bob is not my friend.
Procedural	• A procedure for retrieving an item from a distant location • A procedure for spinning around in a circle • A procedure for stacking a block on top of another one • A procedure for repeatedly asking a question in different ways until an acceptable answer is obtained
Sensory	• The appearance of Bob's face • The specific array of objects on the floor under the table
Episodic	• The series of actions Bill did when he built a tower on the floor yesterday • The episode in which Bill and Bob repeatedly threw a ball back and forth between each other • The series of actions I just took, between getting up from the chair and Bob saying "good"
Attentional	• The set of objects that seem to be important in the context of the game Bob and Bill are playing • The set of words and phrases that are associated with Bob being happy with me while we walk around together
Intentional	• The goal of making Bob say positive things • The goal of making a tower that does not fall down easily • The goal of getting Jim to answer my question

Table 1. Examples of the key knowledge types in the context of virtual agent control

Cognitive Process	Virtual Agent Example
Inference	• Tall thin blocks, when stood upright, are less likely to topple over if placed next to each other • Bob hates cursing, and Jim is Bob's friend, and friends often have similar likes and dislikes, so Jim probably hates cursing
Procedure Learning	• Learning a procedure for crawling on the floor, based on imitation of what others do when they describe themselves as "crawling", plus reinforcement from others when they find one's imitation accurate • Learning a procedure embodying some combination of functional and visual features that predicts whether some entity is considered a toy or not
Attention allocation	• Pictures of women are associated with Bob's happiness, and Bob's happiness is associated with getting reward, therefore pictures of women are associated with getting reward • Asking for help is surprisingly often a precursor to getting reward when Jane is around; so when a reward is gotten when Jane is around, a little extra attention should be given to ongoing improvement of the processes that help in the mechanics of asking for help
Goal refinement	• The goal of making Jim happy, seems to often be achieved by the goal of creating sculptures Jim likes, and Jim likes complicated sculptures; thus I adopt the goal of creating complicated sculptures when Jim is around
Declarative pattern mining	• Tall thin blocks, when stood upright, are likely to topple over
Sensory pattern recognition	• When Jim builds a castle out of blocks, he identifies some portions of the castle as "towers" and others as "walls"; it's necessary to visually identify which portions of each castle correspond to these descriptors • It's also necessary to visually identify the castle as a whole versus the table, floor or other base it's resting on
Simulation	• Using an internal simulation world to experiment with building various towers rapidly, at a pace faster than is possible in the online simulation world where humans participate • Using an internal simulation world containing a simulation of Bob and Jim, to simulate what Bob will know about what you're doing if you hide behind Jim and build a tower of blocks
Concept creation	• The concept of an unstable structure • The concept of an irritable person • The concept of a happy occasion
Map formation	• The set of all knowledge items associated with Bob being in a good mood (which may then be used to form a new concept) • The set of all knowledge items associated with (running, walking or crawling) races

Table 2. Examples of the key cognitive processes in the context of virtual agent control

In general, the cognitive schematic 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

... or, to put it less technically:

1. Evaluating conjectured relationships between procedures, contexts and goals ("analysis")
2. Conceiving novel possible relationships between procedures, contexts and goals ("synthesis"); and when necessary creating new procedures and contexts, via leveraging prior knowledge or as a last resort via trial and error experimentation

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

- Procedural knowledge, and procedural learning methods can be useful for choosing P, given fixed C and G. Simulation may also be useful, via creating a simulation embodying C and seeing which P lead to the simulated achievement of G
- Declarative knowledge, and associated knowledge creation methods, can be useful for choosing C, given fixed P and G (also incorporating sensory and episodic knowledge as useful). Simulation may also be used for this purpose.
- Inference, acting on declarative knowledge, can be useful for choosing G, given fixed P and C. Simulation may also be used for this purpose.
- Goal refinement is used to create new subgoals G to sit on the right hand side of instances of the cognitive schematic
- Concept formation and map formation are useful for choosing G and for fueling goal refinement, but especially for choosing C (via providing new candidates for C). They can also be useful for choosing P, via a process called "predicate schematization" that turns logical predicates (declarative knowledge) into procedures.

On the other hand, where analysis is concerned:

- Inference, acting on 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. Simulation may also be used: multiple simulations may be run, and statistics may be captured therefrom.
- Procedural knowledge, mapped into declarative knowledge and then acted on by inference, 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
- Inference, acting on declarative or 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; and similarly 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

7

How ---> Helps | |/	Map formation	Goal system	Simulation	Sensorimotor pattern recognition
Uncertain inference	Creates new concepts and relationships, enabling briefer useful inference trails	Goal refinement enables more careful goal-based inference pruning	- Simulations provide a method of testing speculative inferential conclusions - Simulations suggest hypotheses to be explored via inference	Creates new concepts and relationships, enabling briefer useful inference trails
Supervised procedure learning	Creates new procedures to be used as modules in candidate procedures	Goal refinement allows more precise definition of fitness functions, making procedure learning's job easier	Simulation provides a method of "fitness estimation" allowing inexpensive testing of candidate procedures	Extraction of sensorimotor patterns allows creation of abstracted fitness functions for (inferentially and simulatively) evaluating procedures guiding real-world actions
Attention allocation	Creates new concepts grouping "attentionally related" memory items, enabling AA to find subtler attentional patterns involving these nodes	Goal refinement allows more accurately goal-driven allocation of attention	Simulation provides data for attention allocation -- allowing attentional information to be extracted from co-occurences observed in simulation	Creates concepts grouping "attentionally related" memory items, enabling AA to find subtler attentional patterns involving these nodes
Concept creation	Creates new concepts to be fed into other concept creation mechanisms	Goal refinement provides more precise definition of criteria via which new concepts are created	Utility of concepts may be assessed via creating simulated entities embodying the new concepts and seeing what they lead to in simulation	Creates new concepts to be fed into other concept creation mechanisms

How ---> Helps | |/	Uncertain inference	Supervised procedure learning	Attention allocation	Concept creation
Uncertain inference	NA	When inference gets stuck in an inference trail, it can ask procedure learning to learn new patterns regarding concepts in the inference trail (if there is adequate data regarding the concepts)	Importance levels allow pruning of inference trees	Provides new concepts, allowing briefer useful inference trails
Supervised procedure learning	Inference can be used to allow prior experience to guide each instance of procedure learning.	NA	Importance levels may be used to bias choices made in the course of procedure learning (e.g. in OCP, in the fitness evaluation and representation-building phases of MOSES)	Provides new concepts, allowing compacter programs using new concepts in various roles
Attention allocation	Enables inference of new HebbianLinks and HebbianPredicates from existing ones	Procedure learning can recognize patterns in historical system activity, which are then used to build concepts and relationships guiding attention allocation	NA	Combination of concepts formed via map formation, may lead to new concepts that even better direct attention
Concept creation	Allows inferential assessment of the value of new concepts	Procedure learning can be used to search for high-quality blends of existing concepts (using e.g. inferential and attentional knowledge in the fitness functions)	Allows assessment of the value of new concepts based on historical attentional knowledge	NA

How ---> Helps | \\\|/	Uncertain inference	Supervised procedure learning	Attention allocation	Concept creation
Map formation	Speculative inference can help map formation guess which maps to hunt for	Procedure learning can be used to search for maps that are more complex than mere "co-occurrence"	Attention allocation provides the raw data for map formation	*No significant direct synergy*
Goal system	Inference can carry out goal refinement	*No significant direct synergy*	Flow of importance among subgoals determines which subgoals get used, versus being forgotten	Concept creation can be used to provide raw data for goal refinement (e.g. a new subgoal that blends two others)
Simulation	In order to provide data for setting up simulations, inference will often be needed	*No significant direct synergy*	Attention allocation tells which portions of a simulation need to be run in more detail	*No significant direct synergy*
Sensorimotor pattern recognition	Speculative inference helps fill in gaps in sensory data	Procedure learning can be used to find subtle patterns in sensorimotor data	Attention allocation guides pattern recognition via indicating which sensorimotor stimuli and patterns tend to be associatively linked	New concepts may be created that then are found to serve as significant patterns in sensorimotor data

How ---> Helps | \\\|/	Map formation	Goal system	Simulation	Sensorimotor pattern recognition
Map formation	NA	Map formation may focus on finding maps related to subgoals, and good subgoal refinement helps here	*No significant direct synergy*	*No significant direct synergy*
Goal system	Concepts formed from maps may be useful raw material for forming subgoals	NA	*No significant direct synergy*	*No significant direct synergy*
Simulation	*No significant direct synergy*	*No significant direct synergy*	NA	Presence of recognized sensorimotor patterns may be used to judge whether a simulation is sufficiently accurate
Sensorimotor pattern recognition	Concepts formed from maps may usefully guide sensorimotor pattern search	Directing pattern search toward patterns pertinent to subgoals, may make the task far easier	Patterns recognized in simulations may then be checked for presence in real sensorimotor data	NA

Table 3. Synergies between the cognitive processes shown in Figure 1.

9

Map formation and concept creation can be useful indirectly in calculating these probability estimates, via providing new concepts that can be used to make useful inference trails more compact and hence easier to construct.

The key role of attentional knowledge in the overall functioning of intelligent systems as described by CST must be emphasized. 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.

Finally, one way to see the essential role of synergy in intelligence as modeled by CST, is to observe that 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. This requires either a procedure learning method that works interactively with a declarative-knowledge-focused concept learning or reasoning method; or a declarative learning method that works interactively with a procedure learning method.

4. ENUMERATION OF CRITICAL SYNERGIES

Referring back to Figure 1, and summarizing many of the ideas in the previous section, Table 1 enumerates a number of specific ways in which the cognitive processes mentioned in the Figure may synergize with one another, potentially achieving dramatically greater efficiency than would be possible on their own.

Of course, realizing these synergies on the practical algorithmic level will require significant inventiveness and may be approached in many different ways. The primary approach we have pursued involves the OpenCogPrime software design, which introduces specific algorithms for each of the capabilities mentioned in Figure 1, together with specific mechanisms for realizing the synergies in Table 1. The specifics of how OpenCogPrime manifests these synergies are discussed further in [7].

References

[1] Solomonoff, Ray. 1964. "A Formal Theory of Inductive Inference, Part I"
Information and Control, Vol 7, No. 1 pp 1-22, March 1964.
[2] Solomonoff, Ray. 1964. "A Formal Theory of Inductive Inference, Part II"
Information and Control, Vol 7, No. 2 pp 224-254, June 1964
[3] Schmidhuber, J. (2006). Gödel machines: Fully Self-Referential Optimal
Universal Self-Improvers. In B. Goertzel and C. Pennachin, eds.: Artificial
General Intelligence, p. 119-226, 2006.
[4] Hutter, Marcus (2005). Universal Artificial Intelligence: Sequential Decisions
based on Algorithmic Probability. Springer, 2005
[5] Goertzel, Ben (2009). The Embodied Communication Prior. Dynamical Psychology
[6] Goertzel, Ben (2009). OpenCog Prime: A Cognitive Synergy Based Architecture for General Intelligence. Submitted for publication
[7] Goertzel, Ben (2006). The Hidden Pattern. Brown Walker
[8] Goertzel, Ben and Stephan Vladimir Bugaj (2009). AGI Preschool. Proceedings of AGI-09, Atlantis Press.