

The Humanoid Robot Cog

by [Naveed Ahmad](#)

Introduction

Rodney Brooks, the inventor of COG and the director of the MIT Artificial Intelligence lab writes [\[5\]](#):

"As I pondered [this] and thought about HAL, I decided to try to build the first serious attempt at a robot with human-level capabilities, the first serious attempt at a HAL-class being."

In 1993, after years of research on behavior-based insect robots, Brooks and his team at MIT started to construct a robot shaped like a human. They named it COG, an abbreviation for "cognition," and also the tooth of a gear. COG was designed and built to emulate human thought processes and experience the world as a human. Brooks and his team further assumed that people would find it easier to interact with a robot and aid the robot in its learning process when the robot could respond in a somewhat human way. Consequently, the machine should have limbs, sensory organs, and a physical resemblance to humans. Unlike other artificial intelligence systems, like medical expert systems, COG was meant to test theories of human cognition and developmental psychology.



Figure 1: Humanoid Robot COG.

What Does COG Look Like?

COG had to be developed physically so it could encounter (rather than simulate) the same environmental and physical constraints as adult humans. Furthermore, human thought and representation is related to the human physical form. Human like intelligence can be only possible with human like form. However, COG does not have legs; from the "waist" down it is built on an immovable stand. It does, however, have a head, torso, and arms. The head has a vision system with four cameras mounted in pairs with two **DOF** (degrees of freedom). Degree of freedom denotes each plane a robot can move in. Each eye is composed of a pair of cameras, one for wide angle and the other for telescopic view. The robot has three rate gyroscopes and two linear accelerometers mounted in the head to mimic the human vestibular system. Two microphones are mounted on the head for the auditory system. Feedback about the motor system is provided by the sensors located at the joints to give feedback about the state of each joint. COG has a total of 22 mechanical DOF, six DOF arms, a torso with a two DOF waist, one DOF torso twist, a four DOF neck, and three DOF eyes. Each DOF in the arm is powered by a DC electric motor through series springs, which provide accurate torque feedback. Changing the equilibrium positions of the joints, not by explicit motor angle commands, determines the position of arms. The spring like arm allows it to move like its human counterpart. COG has a heterogeneous network of processors, ranging from small microcontrollers at joint level to audio and visual processors. The original brain was a network of 16 MHz Motorola 68332 microcontrollers connected through a dual-port RAM, and each node ran a subset of multi-threaded LISP. The current system is a network of PCs running a UNIX real-time operating system connected by 100VG Ethernet. The old and new systems communicate through shared memory interface cards [2].

A Tool to Understand Humans

The most important aspect of humans is their capability to learn and adapt. The two broad categories of human learning are interaction with the physical world and social interaction with fellow human beings. An infant learns to coordinate the movement of its limbs from the simultaneous feedback from its senses. He/she learns the laws of nature from continuous experimentation, and remembers the cause and effect relationships through its limbs and senses. An infant learns to walk after undergoing the stages of kicking, crawling, and stumbling - learning simple behaviors before the hard ones. The other aspect of learning is social interaction, initially with parents, and subsequently through teachers, friends, peers, and others. The use of language, expressions, and emotions make communication possible. All learning is incremental - progressing from simple to more complicated tasks, skills, and concepts. Learning is also cumulative, as previous knowledge serves as the foundation for subsequent knowledge. One way to know if we understand humans is by building one. The bird was both the inspiration and model. After centuries of experimentation in making flying machines led to better understanding of flight, and eventually to the invention of the airplane, similarly COG is an attempt to model humans by constructing a physical body resembling a human and implementing the brain software that imitates human cognition. Perhaps in this way we can better understand human beings and eventually learn more efficient ways in which to teach human children.

Inspiration from the Brain

Efforts to implement COG's mind are inspired from studies in neuroscience. One such example is how COG is taught to orient its head to a moving or noisy stimulus and is inspired from the studies of the Superior Colliculus. The superior Colliculus is the part of the human brain that specializes in integrating sensory information and orienting the sensory organs such as the eye, neck, and ears toward the source of the sensory input. The Superior Colliculus has been found to be organized in layers of topographically arranged maps, where each map is an arrangement of neurons, and the maps are sensitive to certain sensory stimuli. There are also motor maps which upon being stimulated at certain regions, elicit a movement of an organ. These maps are interconnected. The arrangement of the maps and interconnectivity of the maps is known to change over the period of development of a human. In COG, a map is a two-dimensional array of elements in which each element represents a site in the map. Maps are interconnected in a manner through which activity of a site in one map is transferred to another. [Figure 2](#) shows how activity in the visual space is relayed to the motor map, through a visuo-motor registration map. A registration map interconnects two different maps and relays activity from one to the other. There are also registration maps between two sensory maps, for example the visual and audio space, and learning happens by strengthening the connections between the regions which are strongly correlated (and thus activated simultaneously) and weeding out the wrong connections. Visual activity stimulates certain regions of the visual map, which in turn activates the motor map. The region of the most activity in the motor map determines the motor command to direct the eye motion. The error distance between the center of motion and center of view are used to correct the connections between the maps [\[7\]](#).

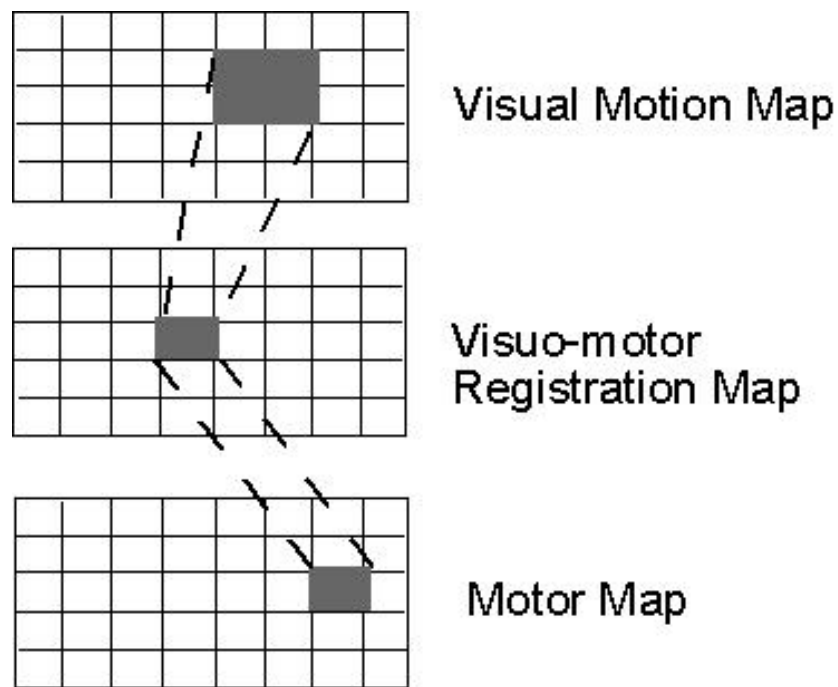


Figure 2: Example of the relationship of COG's visual map to the motor map through the registration map. Activity in a region of the visual map is relayed to the motor map, and the subsequent activity in the motor map determines the commands to the motors of COG's eyes [7].

Incremental Learning

Many AI applications try to take specific approaches to solving specific problems. For example, a typical robot intending to grasp an object would capture video images, process them, make a mathematical model of the coordinates of the object, and send them to a routine which controls the arm's motors in a pre-determined way. As the robot did successfully grasp the object, this solution is effective, however, the robot did not learn the task, and it cannot generalize the concept of *grasping* to another situation. The programmer will always have to devise a solution for every different robotic problem. The researchers at MIT have adopted a developmental approach in which COG learns independently. It learns in a developmental fashion, in a piecemeal way, incrementally and cumulatively. This approach is inspired by the way that infants learn to reach out to and grab objects of interest.

COG learns in two phases. First, COG learns to direct its eyes toward the object by learning a map of the object's coordinates. To learn the motor commands to pan and tilt its gaze, COG repeatedly experiments and readjusts its mapping from the difference of the center to the actual position of the object in the image plane. This is done until COG learns how to align the target object to the center of its image plane. In the second phase, COG experiments with moving its arms and tries to coordinate the ballistic mapping of its arm motors motion to the direction of eyesight (i.e., the direction of sight of the target object). It first attempts to reach by using an initial mapping configuration. It then determines the error distance of its hand and the target object in the image plane. Since COG has already learned to orient its head toward the target object, it applies that knowledge. It knows how to orient its head and eyes toward its hands and makes a correction to its arm-to-sight mapping. In a later attempt, if the target object is in the same eyesight direction as is the hand, the robot will know how to reach it, since it had made a correction in a previous attempt. The system learns the mapping in a few hours of self-training, taking COG approximately 2000 trials [9].

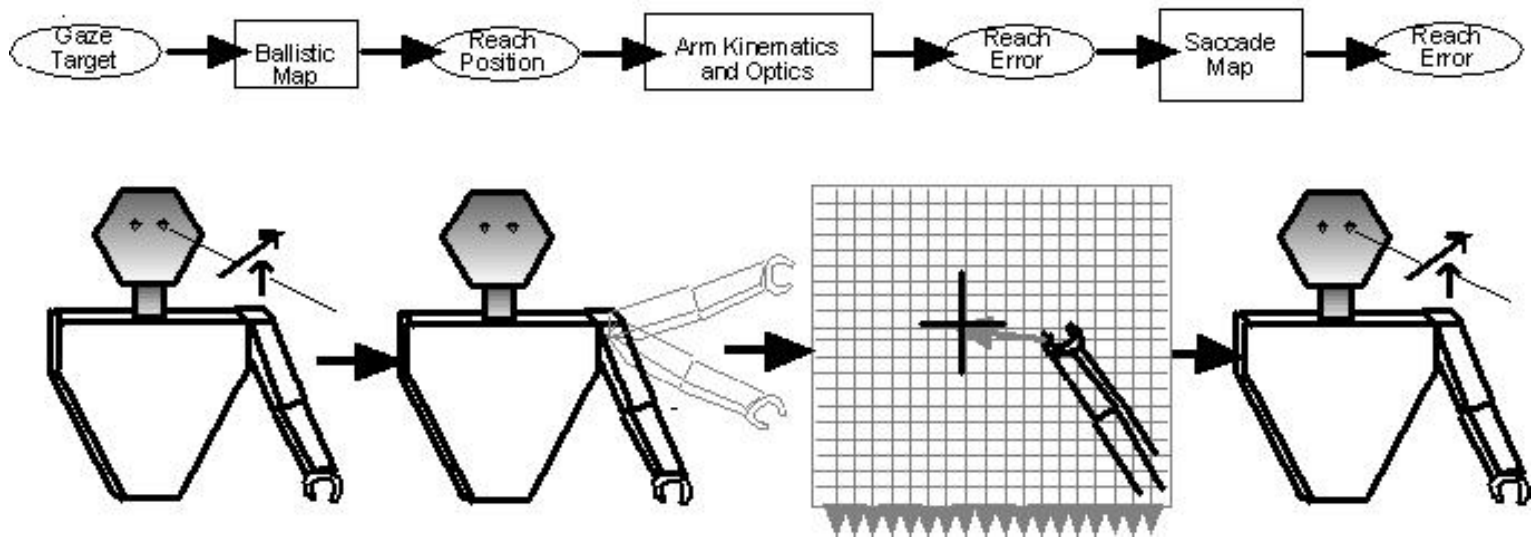


Figure 3: COG orients its head toward the target object, a skill that it mastered in the first phase. It then tries to extend its arm ballistically in the direction of the gaze. If the hand misses the target object, COG uses the knowledge from the first phase. From the position and the gaze direction of the target object for the reach attempt be successful, COG must make a correction in the arm to gaze direction mapping [1].

Social Interaction

Children make many discoveries on their own, just like COG learned to reach an object by learning from its mistakes. But people also learn many things from their parents, teachers, and friends. COG has been programmed to interact socially. The most basic form of communication and social interaction is shared attention (i.e., focus on an object of mutual interest). The stages of learning social skills, in an increasing degree of complexity, are gaze monitoring, following a gaze, following a pointing, and asking for an object by pointing. These skills have been incrementally taught to COG. Learning through a developmental methodology is advantageous because it reduces complex tasks to simpler ones. Complex tasks can reuse the more granular tasks that are easier to learn, and tasks can be learned by gradually increasing their complexity [10]. For example, learning to follow a gaze builds on the skills of holding a gaze. Holding a gaze requires the COG to recognize faces using algorithms which use pre-compiled human face templates. Then, using pre-learned skills, COG orients its head and eyes toward the robot teacher. It then extracts the facial image and locates the eye of the teacher, using the face ratio templates by another learned mapping. Once COG has learned to monitor gaze, the ability to follow a gaze simply builds on the prior skill. Three more additional skills are needed for the more complex skill of gaze following: finding the angle of gaze, extrapolating the angle of gaze to the object, and orienting the motors in the head and eyes to follow the object [11]. This example of shared attention implemented using COG was an example of research into learning social skills in a developmental fashion.

In order for COG to interact socially, it has to show expressions so that the robot teacher knows how to react to the robot. For example, if COG cannot understand the teacher, it could show boredom or puzzlement, hinting that the teacher should try something different. Kismet, COG's stand-alone expressive head, is a relative of COG. Kismet can convey fear, boredom, anger, and other emotions by using its ears, eyelids, mouth, and eyebrows. Kismet has a behavior engine that integrates perceptions, emotions, drives, and behavior. The robot distinguishes between face and non-face stimuli. As long as

Kismet's drives remain in homeostatic range, Kismet displays emotions of happiness and interest. Once the emotions exceed these ranges, it shows expression of fatigue, distress, or fear, depending on the emotional state of Kismet. For example, if the robot is left alone and under-stimulated, it shows expression of sadness meaning that the care giver should play with it. If the robot is overexposed, it may show signs of disgust, meaning that the caregiver should slow down.

Learning and Coordination in the Physical World

Brooks believes that human beings use the world to organize and manipulate knowledge; hence there is no need to build elaborate mathematical models of the world or execute heavy-duty computations before acting. An ant does not need to compute a three-dimensional map of its environment before it moves; it can simply start walking and change directions based on real time cues and landmarks. The robot's behavior is a direct function of its physical interaction with the world. An example of physical coupling is the implementation of COG's arms. The wrists, elbows, and shoulders are driven by force oscillators using proprioceptive information. There is no central controller or modeling of the arms. The complete behavior of the arms is the sum of the behaviors of all the joints responding to the environment [12]. COG's arms can display complex behavior, like balancing a Slinky in its two hands and playing a drum by processing real-time feedback of the sensors, and changing the equilibrium of the oscillators of the arm joints. This movement is produced without elaborate software models of the world, to control the arm or the environment.

The senses are not used singularly by humans, rather they complement each other resulting in an integration of information from multiple senses simultaneously. Just as a bird may chirp to help it identify whether the object overhead is a bird or an airplane, humans may employ lip-reading techniques to assist in listening. COG has learned to use visual information to train its auditory localization. The relationship of a visual movement to the direction of the sound is used to train visual to auditory mapping. Once COG has learned this mapping, it can orient its head toward the source of the sound [8]. COG also mimics human Vestibular Ocular Reflex (VOR). VOR is a reflex that stabilizes an image while the head is moving by turning the eyes in the opposite direction of the direction of head movement. COG learns to compensate its camera motion by feedback from rate of change of the gyroscopes mounted in its head. Relating information from different senses improves the performance of COG and requires less computation than relying on sensory input in isolation. The way that COG learned to reach an object is an example of COG interacting with the physical world, where touching the target object with its arms become a part of COG's physical experience. COG learns to coordinate the sensory information from its vision system with its motors and arms. COG does not construct elaborate models of the world, but simply learns the correlations between the sensory input and its arm's actions.

Not Human as Yet

COG is an ongoing research project to understand and emulate human behavior and psychology. Although the research is only the tip of the iceberg, the goal is to make COG act human. One of the main issues with COG is to make its subsystems and behavior coherent. COG still needs an elaborate motivational model so that it can select between its behaviors. For example, if two objects of interest are in view, COG should be able to decide which object to reach for and which one to ignore based on its goals in focus. Currently COG's behaviors are designed independently and require all of the resources of COG. COG also has far fewer tactile sensors than the human nervous system. It also has sensors to feel

the motion of some joints, but this information has not been used much. The information from the force sensors at each of the joints has not been used on other subsystems except for the direct feedback control of the arms. COG does not yet have the ability to taste and smell. Even though COG has a sophisticated image processing subsystem, it is still behind the capabilities of humans. Although COG can decipher faces and locate motion and colors, it cannot recognize faces in real time. Memory and experience play a key role in human cognition. COG does not experience time; it lives in an eternal present. COG cannot store long-term memories in chronological order; its memory is limited to particular experiments. The challenge is how to relate the static data structures and computational models of memory to the flow of time. These are some of the shortcomings that indicate the future directions of research [2].

Conclusion

Humans are the most complex creatures on earth. Robots still lag behind in terms of cognitive complexity and flexibility. But as theories of human psychology improve and are implemented on human-like robots, COG will evolve from an infant to an adult. Advances in biotechnology may give robots a more biological form, and a greater physical resemblance to humans. COG is not the only humanoid robot. Honda's Asimo, Sony's SDR-4X, and Kitano Symbiotic Systems' baby robot PINO are some of the other humanoid robots being developed around the world. To date, these are not commercially available. Toy robots like Sony's entertainment dog Aibo, however are already on the market. Aibo is an autonomous pet dog robot which inhabits a home, sings, dances, reads emails, re-charges, and learns from its owner. Natural progression in technology ensures that humanoid robots will soon follow commercially available pet robots. We may just be on the verge of a robotic revolution in which robots and intelligent autonomous machines become a common part of our daily lives. The day when we start to communicate with them in human language, teach them our daily chores, and share our responsibilities will be when we have achieved the ultimate science fiction goal of making human-like robots.

References

1

Adams, B., Breazeal, C., Brooks, R., and Scassellati B. Humanoid Robots: A New Kind of Tool *IEEE Intelligent Systems*, July-August 2000, p. 13.

2

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., and Williamson, M. The Cog Project: Building a Humanoid Robot. In *Computation for Metaphors, Analogy and Agents*, Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, Springer-Verlag, 1998. pp. 8-13.

3

Brooks, R. A., Breazeal (Ferrell) C., Irie, R., Kemp, C. C., Marjanovic, M, Scassellati, B., and Williamson, M. Alternate Essences of Intelligence. *AAAI-98*, 1998.

4

Brooks, R. A. *Cambrian Intelligence, The Early History of the New AI*. Cambridge, MA: The MIT Press, 1999, pp. 8-9.

5

Brooks, R. A. *Flesh and Machines, How Robots Will Change Us*. New York: Pantheon Books, 2002, p. 65.

6

The Cog Shop, < <http://www.ai.mit.edu/projects/humanoid-robotics-group/cog/cog>.

[html](#)> (24 May 2003).

7

Ferrell, C. Orientation Behavior Using Registered Topographic Maps. Presented at *SAB 96*, Cape Cod, MA, 1996, p. 5.

8

Irie, R. E. Robust Sound Localization: An Application of an Auditory Perception System for a Humanoid Robot. S.M. Thesis, M.I.T. Department of Electrical Engineering and Computer Science, 1995.

9

Marjanovic, M., Scassellati, B., and Williamson, M. Self-Taught Visually-Guided Pointing for a Humanoid Robot. In *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior (SAB-96)*, September 1996.

10

Scassellati, B. Building Behaviors Developmentally: A New Formalism. In *Proceedings of the 1998 AAAI Spring Symposium on Integrating Robotics Research*, January 1998.

11

Scassellati, B. Imitation and Mechanisms of Joint Attention: A Developmental Structure for Building Social Skills on a Humanoid Robot. In C. Nehaniv, ed., *Computation for Metaphors, Analogy and Agents*. Vol.. 1562 of Springer Lecture Notes in Artificial Intelligence, Springer-Verlag, 1998.

12

Williamson, M. Exploiting Natural Dynamics in Robot Control. In *Proceedings Fourth European Meeting on Cybernetics and Systems Research (EMCSR '98)*, 1998.

Biography

Naveed Ahmad (navacron@hotmail.com) is currently pursuing a Master's degree in Computer Science from the University of Illinois Urbana Champaign. He earned his Bachelor's in Computer Science from Lahore University of Management Sciences. As an undergraduate student, he was involved in research on a cricket expert system. The research project won the Ninth All Pakistan Software Competition. He is a robot hobbyist, and has a keen research interest in AI and making robots act like animals and humans.