SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY GAUTAM BUDDHA UNIVERSITY, GREATER NOIDA, 201312, U. P., (INDIA)

Candidate's Declaration

We, hereby, certify that the work embodied in this **seminar report entitled "Artificial Intelligence"**, submitted in partial fulfilment of the requirements for the award of the degree of **M.Sc. Computer Science**, to the School of Information and Communication Technology, Gautam Buddha University, Greater Noida, is an authentic record of our own research and understanding carried out under the supervision of **Prof. Prakash Kumar Saraswat**, School of ICT. The content presented in this report has not been submitted to any other University or Institute for the award of any other degree or diploma.

NAME	ROLL NO	SIGNATURE
NISHANT GOEL	235/PMD/001	

Supervisor's Certification

This is to certify that the above statement made by the candidates is correct to the best of my knowledge

and belief. However, responsibility for any plagiarism-related issue solely stands with the students.

Signature of the Supervisor:
Name with Designation: **Prof. Prakash Kumar Saraswat** (Supervisor)
Date:
Place: Greater Noida

Acknowledgement

We, **Nishant Goel (235/PMD/001)**, would like to express our heartfelt gratitude to all those who have contributed directly or indirectly to the successful completion of this **seminar report on "Artificial Intelligence."**

First and foremost, we extend our sincere thanks to our seminar guide, **Prof. Prakash Kumar Saraswat**, Assistant Professor (IT), **Gautam Buddha University**, for his invaluable guidance, consistent support, and encouragement throughout the preparation and presentation of this seminar. His insights and expertise were instrumental in shaping our understanding of the topic and ensuring the quality of this report.

We also wish to acknowledge the faculty members of the School of Information and Communication Technology, Gautam Buddha University, for their support and for providing us with a conducive academic environment and access to relevant resources.

Our deepest appreciation goes to our families for their unconditional support, patience, and encouragement throughout the duration of this work.

Lastly, we are thankful to our friends and classmates for their constructive feedback, continuous motivation, and moral support during the entire seminar preparation.

List Of Topics

Abstract

Artificial Intelligence (AI) research has long struggled with the challenge of knowledge representation—how to formally capture, store, and manipulate information about the world in a way that machines can reason about. Traditionally, AI systems have relied heavily on complex symbolic representations and centralized planning modules. However, this approach has often led to brittle systems that fail to perform effectively in real-world, dynamic environments due to the unpredictability and richness of sensory inputs and the difficulty of modelling all possible scenarios.

In contrast, recent advancements advocate for a more incremental and embodied approach to intelligence, where the focus is on direct interaction with the real world through continuous perception and action. Rather than constructing high-level internal models, intelligent behaviour emerges from a tight coupling between sensing and acting. In this paradigm, the need for complex internal representations diminishes significantly.

In this paper, we present our approach to building intelligent autonomous agents—referred to as “Creatures”—incrementally, by embedding them directly in real-world environments. Our methodology does not decompose the intelligent system into discrete, sequential modules that communicate via symbolic representations. Instead, the system is designed as a collection of independent, parallel activity-producing modules, each responsible for a particular behaviour or response, and all interfacing directly with the environment.

This architecture eliminates the traditional distinction between “central” and “peripheral” systems. Every module in our system acts both as a sensor and an effector—it perceives, processes, and acts in parallel with others. As a result, the system becomes more robust, adaptive, and scalable, and can continue functioning even when some modules fail.

Based on these principles, we have developed a series of mobile robots capable of autonomous operation in standard office environments. These robots require no human supervision and demonstrate effective navigation, obstacle avoidance, and task execution using only direct sensory-motor interactions. The success of these robots validates the effectiveness of the incremental, behaviour-based approach and offers a promising direction for the future of intelligent systems design.

CHAPTER -1

Introduction

Artificial intelligence started as a field whose goal was to replicate human level intelligence in a machine. Early hopes diminished as the magnitude and difficulty of that goal was appreciated. Slow progress was made over the next 25 years in demonstrating isolated aspects of intelligence. Recent work has tended to concentrate on commercial aspects of "intelligent assistants" for human workers. No one talks about replicating the full gamut of human intelligence any more. Instead, we see a retreat

into specialized sub problems, such as ways to represent knowledge, natural language understanding, vision, or even more specialized areas such as truth maintenance systems or plan verification. All the work in these subareas is benchmarked against the sorts of tasks humans do within those areas.

Amongst the dreamers still in the field of AI (those not dreaming about dollars, that is), there is a feeling. That one day all these pieces will all fall into place and we will see "truly" intelligent systems emerge. However, I, and others, believe that human level intelligence is too complex and little understood to be correctly decomposed into the right sub pieces now and that even if we knew the sub pieces, we still wouldn't know the right interfaces between them. Furthermore, we will never understand how to decompose human level intelligence until we have had a lot of practice with simpler level intelligences.

In this paper I therefore argu e for a different approach to creating artificial intelligence:

- We must incrementally build up the capabilities of intelligent systems, having complete systems at each step of the way and thus automatically ensure that the pieces and their interfaces are valid.
- At each step we should build complete intelligent systems that we let loose in the real world with real sensing and real action. Anything less provides a candidate with which we can delude ourselves.

We have been following this approach and have built a series of autonomous mobile robots. We have reached an unexpected conclusion (C) and have a rather radical hypothesis (H).

(C) When we examine very simple level intelligence, we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model.

(H) Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems.

Representation has been the central issue in artificial intelligence work over the last 15 years only because it has provided an interface between otherwise isolated modules and conference papers.

CHAPTER -2

The Evolution of Intelligence

The very existence of human beings serves as definitive proof that the emergence of intelligent entities is not only possible but naturally achievable. Furthermore, numerous animal species exhibit varying degrees of intelligence, although debates around the nature and definition of intelligence remain ongoing. Nevertheless, what is beyond dispute is that intelligence—however we define it—has evolved over Earth's extensive 4.6-billion-year history.

To appreciate the developmental trajectory of intelligence, it is instructive to reflect on the timeline of biological evolution. Life began with simple single-celled organisms emerging from the primordial soup approximately 3.5 billion years ago. One billion years later, photosynthetic plants evolved, marking a significant advancement in energy processing. After nearly another billion and a half years, around 550 million years ago, the first vertebrates and fish appeared. Insects followed around 450 million years ago, and from that point forward, evolutionary advancements accelerated.

Reptiles emerged around 370 million years ago, followed by the rise of dinosaurs at about 330 million years ago and mammals at 250 million years ago. The evolutionary branch that led to primates emerged roughly 120 million years ago, and the ancestors of the great apes appeared only around 18 million years ago. Homo sapiens, in a form recognizable today, arrived just 2.5 million years ago. Remarkably, agriculture was developed a mere 10,000 years ago, writing less than 5,000 years ago, and expert knowledge systems only within the past few centuries.

This timeline strongly suggests that the foundations of intelligence—such as problem-solving, communication, symbolic reasoning, and knowledge application—are relatively recent and possibly simpler developments in comparison to the long and complex evolution of sensory processing, motor coordination, and survival behaviour. These latter faculties—essential to life in dynamic environments—appear to form the core upon which higher intelligence is built. Evolution has spent the majority of its time refining mobility, perception (especially vision), and behaviour geared toward survival and reproduction, indicating the inherent complexity and importance of these capabilities.

This observation aligns with the arguments of Hans Moravec and other roboticists, who posit that true intelligence must emerge from grounded experience in the physical world. According to this perspective, it is not enough to focus on high-level reasoning or symbolic processing in isolation. Instead, artificial intelligence must begin with the capacity to perceive and act within the world—just as biological organisms do.

A powerful analogy can be drawn from the history of artificial flight. Imagine a group of flight enthusiasts and engineers in the 1890s, deeply engaged in the pursuit of powered flight. Suppose they are transported briefly into the 1980s and allowed to spend time inside a Boeing 747 mid-flight. Overwhelmed by the marvel of air travel, they return to their time filled with inspiration and confidence that flight on such a scale is achievable. However, without having seen the aircraft's engines or underlying mechanics, they focus on replicating superficial features—padded seats, windows, and unfamiliar materials like plastic. A few among them may have glimpsed part of the engine and become fascinated by its complexity, but their understanding remains fragmented.

This thought experiment warns us not to mistake end products of intelligence for its foundational mechanisms. Just as the essence of flight lies in the physics of aerodynamics and propulsion—not in passenger comfort—the essence of intelligence lies not in expert systems or symbolic manipulation, but in grounded interaction with the world. Attempts to recreate intelligence must, therefore, begin not with mimicking high-level human cognition, but with developing systems capable of autonomous behavior, perception, and adaptation in real environments.

CHAPTER -3

Abstraction as a dangerous weapon

Artificial intelligence researchers are fond of pointing out that AI is often denied its rightful successes. The popular story goes that when nobody has any good idea of how to solve a particular sort of problem (e.g. playing chess) it is known as an AI problem. When
an algorithm developed by AI researchers successfully tackles such a problem, however, AI detractors claim that since the problem was solvable by an algorithm, it wasn't really an AI problem after all. Thus AI never has any successes. But have you ever heard of an AI failure?
I claim that AI researchers are guilty of the same (self) deception. They partition the problems they work on into two components. The AI component, which they solve, and the non-AI component which, they don't solve. Typically, AI "succeeds" by defining the parts of the problem that are unsolved as not AI.

The principal mechanism for this partitioning is abstraction. Its application is usually considered part of good science, not, as it is in fact used in AI, as a mechanism for self-delusion. In AI, abstraction is usually used to factor out all aspects of perception and motor skills. I argue below that these are the hard problems solved by intelligent systems, and further that the shape of solutions to these problems constrains greatly the correct solutions of the small pieces of intelligence which remain.

Early work in AI concentrated on games, geometrical problems, symbolic algebra, theorem proving, and other formal systems. In the late sixties and early seventies the blocks world became a popular domain for AI research. It had a uniform and simple semantics. The key to success was to represent the state of the world completely and explicitly. Search techniques could then be used for planning within this well-understood world. Learning could also be done within the block’s world; there were only a few simple concepts worth learning and they could be captured by enumerating the set of Sub expressions which must be contained in any formal description of a world including an instance of the concept. The blocks world was even used for vision research and mobile robotics, as it provided strong constraints on the perceptual processing necessary [12].

Eventually criticism surfaced that the blocks world was a "toy world" and that within it there were simple special purpose solutions to what should be considered more general problems. At the same time there was a funding crisis within AI (both in the US and the UK, the two most active places for AI research at the time). AI researchers found themselves forced to become relevant. They moved into more complex domains, such as trip planning, going to a restaurant, medical diagnosis, etc. Soon there was a new slogan: "Good representation is the key to AI" (e.g. conceptually efficient programs in [2]). The idea was that by representing only the pertinent facts explicitly, the semantics of a world (which on the surface was quite complex) were reduced to a simple closed system once again.

Abstraction to only the relevant details thus simplified the problems.
Consider a chair for example. While the following two characterizations are true:
(CAN (SIT-ON PERSON CHAIR)), (STAND-ON PERSON CHAIR)), there is much more to the concept of a chair. Chairs have some flat (maybe) sitting place, with perhaps a back support. They have a range of possible sizes, requirements on strength, and- a range of possibilities in shape. They often have covering material, unless they are made of wood, metal, or plastic. They sometimes are soft in particular places.

They can come from a range of possible styles. The concept of what is a chair is hard to characterize simply. There is certainly no AI vision program which can find arbitrary chairs in arbitrary images; they can at best find one chair in carefully selected images.

This characterization, however, is perhaps the correct AI representation of solving certain problems; e.g., a person sitting on a chair in a room is hungry and can see a banana hanging from the ceiling just out of reach. Such problems are never posed to AI systems by showing them a photo of the scene. A person (even a young child) can make the right interpretation of the photo and suggest a plan of action. For AI planning systems however, the experimenter is required to abstract away most of the details to form a simple description in terms of atomic concepts such as PERSON, CHAIR and BANANAS.

But this abstraction is the essence of intelligence and the hard part of the problems being solved. Under the current scheme the abstraction is done by the researchers leaving little for the AI programs to do but search. A truly intelligent program would study the photograph, perform the abstraction and solve the problem. The only input to most AI programs is a restricted set of simple assertions deduced from the real data by humans. The problems of recognition, spatial understanding, and dealing with sensor noise, partial models, etc. are all ignored. These problems are relegated to the realm of input black boxes. Psychophysical evidence suggests they are all intimately tied up with the representation of the world used by an intelligent system. There is no clean division between perceptions (abstraction) and reasoning in the real. world. The brittleness of current AI systems attests to this fact.
For example, MYCIN [13] is an expert at diagnosing human bacterial infections, but it really has no model of what a human (or any living creature) is or how they work, or what are plausible things to happen to a human. If told that the aorta is ruptured and the patient is losing blood at the rate of a pint every minute, MYCIN will try to find a bacterial cause of the problem.

Thus, because we still perform all the abstractions for our programs, most AI work is still done in the blocks world. Now the blocks have slightly different shapes and colors, but their underlying semantics have not changed greatly. It could be argued that performing this abstraction (perception) for AI programs is merely the normal reductionist use of abstraction common in all good science. The abstraction reduces the input data so that the program experiences the same perceptual world (Merkwelt in [15]) as humans. Other (vision) researchers will independently fill in the details at some other time and place. I object to this on two grounds. First, as Uexküll and others have pointed out, each animal species, and clearly each robot species with their own distinctly non-human sensor suites, will have their own different Merkwelt.

Second, the Merkwelt we humans provide our programs is based on our own introspection. It is by no means clear that such a Merkwelt *is* anything like what we actually use internally—it could just as easily be an output coding for communication purposes (e.g., most humans go through life never realizing, they have a large blind spot almost in the center of their visual fields).
The first objection warns of the danger that reasoning strategies developed for the human-assumed Merkwelt may not be valid when real sensors and perception processing is used. The second objection says that even with human sensors and perception the

Merkwelt may not be anything like that used by humans. In fact, it may be the case that our introspective descriptions of our internal representations are completely misleading and quite different from what we really use.

3.1. A continuing story

Meanwhile our friends in the 1890s are busy at work on their AF machine. They have come to agree that the project is too big to be worked on as a single entity and that they will need to become specialists in different areas. After all, they had asked questions of fellow passengers on their flight and discovered that the Boeing Co. employed over 6000 people to build such an airplane.

Everyone is busy but there is not a lot of communication between the groups. The people making the passenger seats used the finest solid steel available as the framework. There was some muttering that perhaps they should use tubular steel to save weight, but the consensus was that if such an obviously big and heavy airplane could fly then clearly there was no problem with weight. On their observation flight none of the original group managed to get a glimpse of the driver's seat, but they have done some hard thinking and think they have established the major constraints on what should be there and how it should work. The pilot, as he will be called, sits in a seat above a glass floor so that he can see the ground below so he will know where to land. There are some side mirrors so he can watch behind for other approaching airplanes. His controls consist of a foot pedal to control speed (just as in these newfangled automobiles that are starting to appear), and a steering wheel to turn left and right.

In addition, the wheel stem can be pushed forward and back to make the airplane go up and down. A clever arrangement of pipes measures airspeed of the airplane and displays it on a dial. What more could one want? Oh yes. There's a rather nice setup of louvers in the windows so that the driver can get fresh air without getting the full blast of the wind in his face.An interesting sidelight is that all the researchers have by now abandoned the study of aerodynamics.

Some of them had intensely questioned their fellow passengers on this subject and not one of the modern flyers had known a thing about it. Clearly the AF researchers had previously been wasting their time in its pursuit.

CHAPTER -4

Designing Autonomous Intelligent Creatures

I wish to build completely autonomous mobile agents that co-exist in the world with humans, and are seen by those humans as intelligent beings in their own right. I will call such agents *Creatures.* This is my intellectual motivation. I have no particular interest in demonstrating how human beings work, although humans, like other animals, are interesting objects of study in this endeavour as they are successful autonomous agents. I have no particular interest in applications it seems clear to me that if my goals can be met then the range of applications for such Creatures will be limited only by our (or their) imagination. I have no particular interest in the philosophical implications of Creatures, although clearly there will be significant implications.

Given the caveats of the previous two sections and considering the parable of the AF researchers, I am convinced that I must tread carefully in this endeavour to avoid some nasty pitfalls.
For the moment then, consider the problem of building Creatures as an engineering problem. We will develop an *engineering methodology* for building Creatures.

First, let us consider some of the requirements for our Creatures.
• A Creature must cope appropriately and in a timely fashion with changes in its dynamic environment.
• A Creature should be robust with respect to its environment; minor changes in the properties of the world should not lead to total collapse of the
Creature's behaviour; rather one should expect only a gradual change in capabilities of the creature as the environment changes more and more.
• A Creature should be able to maintain multiple goals and, depending on the circumstances it finds itself in, change which goals it is actively pursuing; thus, it can both adapt to surroundings and capitalize on fortuitous circumstances.
• A Creature should do something in the world; it should have some purpose in being.
Now, let us consider some of the valid engineering approaches to achieving these requirements. As in all engineering endeavours it is necessary to decompose a complex system into parts, build the parts, then interface them into a complete system

4.1 Vision and Motivation

My intellectual motivation lies in the aspiration to build completely autonomous mobile agents—referred to as Creatures—that can coexist naturally in human environments and be perceived by humans as intelligent beings. These Creatures should not be mere tools, but autonomous entities capable of operating independently in dynamic settings.

The pursuit of this goal is not driven by a desire to replicate the workings of the human brain, nor by a primary interest in specific applications or philosophical implications. Rather, the focus is on engineering intelligence, through interaction with the real world, to create agents that can function independently. Applications, while not the immediate concern, will naturally follow if such agents are realized. Their potential will be limited only by our imagination—or perhaps their own.
However, as with any ambitious undertaking, one must tread carefully. Drawing a parallel to the earlier analogy of 19th-century artificial flight (AF) researchers exposed briefly to a modern Boeing 747, we must avoid focusing on superficial aspects of intelligence (symbolic manipulation, high-level logic) while overlooking its true foundations—perception, mobility, and adaptation to real-world dynamics.

4.2 Requirements of Intelligent Creatures

To build such autonomous agents, certain engineering requirements must be met:

- Timely Environmental Response: The creature must be able to sense and respond appropriately to changes in its environment in real time.
- Robustness: It should be resilient to environmental variations. Small changes in external conditions should not lead to total system failure; instead, performance may degrade gracefully.
- Goal Flexibility: It must be capable of pursuing multiple goals and dynamically prioritizing them based on context and opportunity.
- Purposeful Action: It should perform meaningful tasks that demonstrate purpose and intentional behaviour in the real world.

4.3 Approaches to System Decomposition

As in all engineering disciplines, building a complex system such as an intelligent Creature requires decomposition—breaking the system down into manageable components. Two contrasting approaches to this decomposition exist:

4.3.1 Functional Decomposition

The traditional AI methodology is based on functional decomposition. The intelligent system is divided into:

- Perceptual Modules: Convert raw sensor data into symbolic representations.
- Central Symbolic Processor: Performs reasoning, planning, and decision-making.
- Action Modules: Translate symbolic action commands into physical movements.

This approach mirrors classical software architecture. Researchers often specialize in one component (e.g., vision or planning) and make assumptions about the input/output interfaces without integrating into a whole system. As a result, the overall system may be fragmented or even non-functional due to incompatible assumptions across components.

Moreover, subfields such as knowledge representation, learning, planning, and qualitative reasoning have emerged as isolated modules with poorly defined or inconsistent interfaces. This creates significant engineering challenges, especially when one must build the entire chain from perception to action before any testing can occur.

4.3.2 Activity-Based Decomposition

An alternative and more promising methodology are activity-based decomposition, where the intelligent system is divided into independent activity-producing subsystems, also called layers. Each layer is responsible for a specific behaviour and connects sensing directly to action—eliminating the need for symbolic intermediate representations.

This paradigm shift recognizes that intelligence emerges not from a central controller but from the interactions between multiple autonomous behaviour modules, each responsible for a particular goal or survival mechanism.
For example, one basic layer may enable the creature to avoid obstacles, using simple sensor data to steer away from physical objects. This layer functions independently and robustly, forming a fully functional autonomous system. Subsequent layers can then be incrementally added, each introducing more complex behaviour, such as navigating to visible goals or interacting with humans. Each new layer operates in parallel with the lower layers, possibly injecting commands or influencing behaviour, but without overriding or replacing the foundational layers. The original simple behaviours remain active, ensuring robustness and layered intelligence.

In previous experiments ([Brooks, 1986]), the first layer allowed a robot to avoid obstacles. A second layer was later added to guide the robot towards distant visual targets. The higher layer directed motion toward the goal, while the original avoidance layer continued to prevent collisions. These two layers interacted without explicit communication, demonstrating how distributed control and layered behaviour can yield intelligent outcomes without centralized symbolic reasoning.

4.4 Engineering Philosophy

The incremental construction of autonomous agents, starting from minimal behaviours and building up through layers, offers a practical and scalable path toward achieving true intelligence. This approach emphasizes:

- Real-world testing at every stage
- Reusability and robustness of foundational behaviours
- Elimination of rigid symbolic intermediates
- Flexibility to adapt and extend capabilities as needed

Ultimately, this methodology aligns with biological evolution—where intelligence arose not from abstract logic but from adaptive behaviour in dynamic environments. By grounding artificial systems in real-world interaction, we move closer to building truly autonomous, intelligent Creatures.

CHAPTER -5

Who has the representations?

With multiple layers, the notion of perception delivering a description of the world gets blurred even more as the part of the system doing perception is spread out over many pieces which are not particularly connected by data paths or related by function. Certainly, there is no identifiable place where the "output" of perception can be found. Furthermore, totally different sorts of processing of the sensor data proceed independently and in parallel, each affecting the overall system activity through quite different channels of control. In fact, not by design, but rather by observation we note that a common theme in the ways in which our layered and distributed approach helps our Creatures meet our goals is that there is no central representation.

- Low-level simple activities can instil the creature with reactions to dangerous or important changes in its environment. Without complex representations and the need to maintain those representations and reason about them, these reactions can easily be made quick enough to serve their purpose. The key idea is to sense the environment often, and so have an up-to-date idea of what is happening in the world.

- By having multiple parallel activities, and by removing the idea of a central representation, there is less chance that any given change in the class of properties enjoyed by the world can cause total collapse of the system. Rather one might expect that a given change will at most incapacitate some but not all the levels of control. Gradually as a more alien world is entered (alien in the sense that the properties it holds are different from the properties of the world in which the individual layers were debugged), the performance of the creature might continue to degrade. By not trying to have an analogous model of the world, centrally located in the system, we are less likely to have built in a dependence on that model being completely accurate. Rather, individual layers extract only those *aspects* [1] of the world which they find relevant-projections of a representation into a simple subspace, if you like. Changes in the fundamental structure of the world have less chance of being reflected in every one of those projections than they would have of showing up as a difficulty in matching some query to a central single world model.

- Each layer of control can be thought of as having its own implicit purpose (or goal if you insist). Since they are *active* layers, running in parallel and with access to sensors, they can monitor the environment and decide on the appropriateness of their goals. Sometimes goals can be abandoned when circumstances seem unpromising, and other times fortuitous circumstances can be taken advantage of. The key idea here is to be using the world as its own model and to continuously match the preconditions of each goal against the real world. Because there is separate hardware for each layer we can match as many goals as can exist in parallel, and do not pay any price for higher numbers of goals as we would if we tried to add more and more sophistication to a single processor, or even some multiprocessor with a capacity-bounded network.

- The purpose of the creature is implicit in its higher-level purposes, goals or layers. There need be no explicit representation of goals that some central (or distributed) process selects from to decide what. is most appropriate for the creature to do next.

5.1. No representation versus no central representation

Just as there is no central representation there is not even a central system. Each activity producing layer connects perception to action directly. It is only the observer of the creature who imputes a central representation or central control. The creature itself has none; it is a collection of competing behaviours.

Out of the local chaos of their interactions there emerges, in the eye of an observer, a coherent pattern of behaviour. There is no central purposeful locus of control. Minsky gives a similar account of how human behaviour is generated. Note carefully that we are not claiming that chaos
is a necessary ingredient of intelligent behaviour. Indeed, we advocate careful engineering of all the interactions within the system (evolution had the luxury of incredibly long-time scales and enormous
numbers of individual experiments and thus perhaps was able to do without this careful engineering).

We do claim however, that there need be no explicit representation of either the world or the intentions of the system to generate intelligent behaviours for a Creature. Without such explicit representations, and when viewed locally, the interactions may indeed seem chaotic and without purpose.

I claim there is more than this, however. Even at a local, level we do not have traditional AI representations. We never use tokens which have any semantics that can be attached to them. The best that can be said in our implementation is that one number is passed from a process to another. But it is only by looking at the state of both the first and second processes that that number can be given any interpretation at all. An extremist might say that we really do have representations, but that they are just implicit. With an appropriate mapping of the complete system and its state to another domain, we could define a representation that these numbers and topological connections between processes somehow encode. However, we are not happy with calling such things a representation. They differ from standard representations in too many ways.

There are no Variables that need instantiation in reasoning processes. There are no rules which need to be selected through pattern matching. There are no choices to be made. The state of the world largely determines the action of the creature. Simon noted that the complexity of behaviour of a system was not necessarily inherent in the complexity of the creature, but Perhaps in the complexity of the environment. He made this analysis in his description of an Ant wandering the beach, but ignored its implications in the next paragraph when he talked about humans. We hypothesize (following Agre and Chapman) that much of even human level activity is similarly a reflection of the world through very simple mechanisms without detailed representations.

CHAPTER -6

The methodology, in practice

To build robust systems based on activity decomposition, one must follow a careful and disciplined methodology. The first methodological maxim emphasizes the importance of testing intelligent agents—referred to here as *Creatures*—directly in the real world, the same complex and unpredictable environment that humans inhabit. It is tempting to begin development in simplified, artificial environments (e.g., using matte-painted walls, neat right angles, or bright colored blocks as obstacles) with the intention of later transitioning the system to real-world conditions. However, this approach often leads to subtle yet critical dependencies forming between system components and the artificial properties of the simplified environment. These dependencies can become embedded in the assumptions and interfaces of submodules. When the system is eventually exposed to the real world, the assumptions no longer hold, requiring not only major revisions to individual components but potentially a complete rethinking of the system's overall architecture.

Instead of beginning with simplifications, each behavior-based layer must be constructed and tested in the full complexity of the real world. The system must operate in this environment for extended periods, allowing careful observation and thorough debugging of its behavior. As new layers are added incrementally, it becomes essential to isolate bugs. If the lower layer is already stable and verified, then any issues that emerge during integration with a new layer are far easier to localize and fix. The process becomes a controlled experiment with a single variable—the newly added layer.

This methodology was exemplified in the development of four mobile robots at the MIT Artificial Intelligence Laboratory, all of which were designed to operate in unconstrained, dynamic environments such as labs and office corridors. These robots interacted with humans walking by, standing nearby, or even deliberately attempting to confuse them. Each robot functioned autonomously from the moment it was powered on, pursuing goals defined by its layered behavioral architecture. Unlike traditional robots that execute pre-programmed missions or follow explicit plans, these Creatures embody an autonomous, emergent intelligence based on layers of behavior. Although four robots were built, two were identical, resulting in three distinct designs. One robot relied on an offboard LISP machine for computation, two others used onboard combinational logic networks, and one featured a custom parallel processor.

All these robots shared a common architectural foundation known as the subsumption architecture. This approach decomposes control into layers of task-achieving behaviors, with systems incrementally constructed and refined through real-world testing. Each layer consists of a fixed-topology network of simple finite state machines (FSMs). These FSMs have a small number of states, a couple of internal registers and timers, and access to basic computation units capable of tasks like vector summation. The FSMs operate asynchronously, exchanging fixed-length messages—1-bit on the smaller robots, 24-bit on the larger ones—through either virtual or physical wires.

There is no central control mechanism. Instead, each finite state machine responds only to the arrival of messages or the expiration of timers. The machines process messages by either branching conditionally, outputting data, or invoking a simple computational routine. Crucially, there is no access to global data or dynamic communication links, making any form of centralized control impossible. The architecture enforces a strict locality of processing, with each FSM functioning autonomously yet constrained by its hardwired connections.

Layer integration in this architecture is achieved through two key mechanisms: suppression and inhibition. In suppression, a new wire is connected to the input of an existing FSM. When a message arrives on this new wire, it is treated as if it had arrived through the original wire, but it also causes suppression of future messages on the original wire for a predefined duration. In inhibition, the new wire taps into the output of an FSM. A message on this wire simply prevents the FSM from sending its own output for a certain time, without replacing it. These techniques allow higher-level behaviors to modulate lower-level ones without violating the modularity or autonomy of individual FSMs.

To illustrate this architecture in practice, consider a three-layer control system implemented on the first of these robots, which was used for over a year in real-world environments. The robot is equipped with a ring of twelve ultrasonic sonar sensors that fire every second to provide radial depth measurements. These sonar readings are notoriously noisy, suffering from issues like specular reflection and multi-path echoes, particularly at shallow incidence angles.

The lowest layer of control implements collision avoidance. A sonar FSM gathers sensor readings and translates them into a polar map, which is fed to two other FSMs: collide and feelforce. The collide FSM halts the robot if an object is detected directly ahead, while the feelforce FSM calculates a repulsive force based on the inverse-square law using the sonar data. This force vector is passed to the runaway FSM, which thresholds it and forwards it to the turn FSM, reorienting the robot away from obstacles. Simultaneously, the forward FSM propels the robot forward, unless interrupted by a halt command. Together, these FSMs enable the robot to safely navigate a dynamic environment, halting or retreating from perceived threats.

The second layer adds wandering behavior. A wander FSM periodically generates random headings. These are treated by an avoid FSM as attractive forces, which are summed with the repulsive forces from the sonar readings. This composite force determines the robot's heading. Through suppression, the wander layer influences the lower-level avoidance FSMs, nudging the robot in randomly selected directions while still respecting obstacle avoidance. If the avoidance system is actively engaged, wander signals may be ignored, ensuring safety remains paramount.

The third layer introduces exploration. This layer activates when the robot is idle. A whenlook FSM triggers the free space finder FSM (called stereo in the diagram) to search for distant open areas and simultaneously inhibits wandering behavior to maintain observation integrity. When a clear path is found, a command is sent to the pathplan FSM, which selects a heading and sends it to the avoid FSM, integrating goal-directed movement into the control loop. The actual motion of the robot is tracked by the integrate FSM, which feeds position updates back to pathplan. This enables path correction, ensuring that the robot continues toward its goal while dynamically avoiding obstacles.

These behaviors illustrate how layers are constructed independently but interact seamlessly to produce coherent, goal-directed behavior. Importantly, lower-level behaviors do not depend on higher-level ones, preserving robustness and modularity. This design exemplifies the principles of the subsumption architecture: layered, decentralized control, emergent behavior through interaction, and rigorous, real-world incremental development.

CHAPTER -7

Limits to Growth

Since our approach is a performance-based one, it is the performance of the systems we build that must be used to measure their usefulness and to identify their limitations. As of mid-1987, we claim that our robots, using the subsumption architecture to implement complete Creatures, are the most reactive real-time mobile robots in existence. In contrast to most other mobile robots, which are limited to individual experimental runs in static or fully mapped environments, our robots operate completely autonomously in complex, dynamic settings. They begin interacting with the world the moment they are powered on and continue functioning until their batteries are depleted. We believe that their performance level is more comparable to simple insect intelligence than to bacterial-level intelligence. While our goal is to achieve insect-level intelligence within two years—a goal that evolution took three billion years to approach—we recognize the nontrivial nature of that target. This is not a prediction but rather an indication of the scale and complexity of the challenge.

Despite encouraging results so far, our approach faces serious questions. We have ideas and hopes for how to resolve them, but only system performance can ultimately validate our beliefs. Since experimental development takes time, and acknowledging that some of the experiments discussed here have yet to be performed, we aim to outline a plausible path forward. Our intent is not to make definitive claims but to indicate that continued progress toward more intelligent machines is feasible from our current foundation. We believe the layers of activity-producing control—such as those managing mobility, vision, and survival tasks—are necessary building blocks for the emergence of higher-level intelligence akin to human cognition.

One of the most natural and significant questions concerns the limits of our architecture: how many layers of control can be constructed before their interactions become too complex to manage? To date, the highest number of layers we have deployed on a physical robot is three, while in simulation we have tested up to six parallel layers. So far, the methodology of fully debugging each existing activity-producing layer before adding a new one has proven practical. We are currently progressing toward a system on our fourth robot that will incorporate approximately fourteen individual layers.

This fourth robot will exhibit significantly more complex behavior. It uses infrared proximity sensors for local obstacle avoidance and features an onboard manipulator capable of grasping objects from the ground and tabletop surfaces, estimating their weight, and homing in on target objects with onboard depth sensors. A structured light laser scanner is being developed to generate rough depth maps of the forward field of view. The high-level behavior we aim to develop for this Creature involves autonomously navigating office areas, identifying open office doors, entering those rooms, locating and retrieving empty soda cans from cluttered desks, and returning them to a central repository.

To achieve this overall behavior, the robot must integrate numerous simpler, task-specific behaviors. These include: avoiding objects, following walls, recognizing and passing through doorways, aligning with learned landmarks, heading homeward, learning bearings at key landmarks and following them, identifying table-like objects, approaching them, scanning their surfaces for soda can-sized cylinders, positioning the manipulator above detected objects, analyzing their structure via hand-mounted sensors, grasping them if deemed light enough, and depositing them in the repository. Importantly, these tasks do not require coordination by a central controller. Instead, each behavior can be triggered by environmental cues. For example, the grasping behavior will activate only when the manipulator's sensors detect an object of suitable size in an appropriate location. If, upon closer inspection, the object does not resemble a soda can, the grasp reflex is suppressed, and lower-level behaviors prompt the robot to search elsewhere.

Another key question is whether higher-level functions like learning can be achieved within these fixed-topology networks of simple finite state machines. Interestingly, some insects demonstrate a form of learning often referred to as "learning by instinct." For example, honey bees appear to be pre-wired to learn to recognize certain flowers and navigate to and from the hive. Butterflies can also learn to distinguish between different flowers, though in a limited way—when forced to learn about a new type of flower, they often forget the previously known one, suggesting a fixed capacity for information retention.

We have identified ways to construct fixed-topology networks of finite state machines capable of such learning. At present, these networks function as isolated subsystems, with their inputs and outputs yet to be fully integrated into our complete robotic systems. Ironically, this puts us in the very position we earlier criticized in other AI research—developing isolated modules without full integration. We are working to correct this, but progress is slow. Physical experiments with real robots in real environments are inherently difficult and time-consuming. Furthermore, most off-the-shelf hardware and software come with rigid assumptions about their use, making them incompatible with the flexibility our systems demand. Consequently, as of mid-1987, our progress in learning has been delayed by the need to engineer a new video camera and a high-speed, low-power processor to execute our custom vision algorithms at ten frames per second. Each of these components represents a significant engineering challenge, which we are addressing as rapidly as our resources allow.

Of course, theoretical discussion is easy, but real progress depends on empirical performance. Only continued experimentation with physical robots in real-world conditions will reveal the true limitations and potential of our approach. Ultimately, time and experimentation will tell whether our path leads to more intelligent machines.

CHAPTER -8

References

[1] P.E. Agre and D. Chapman, *Unpublished memo*, MIT Artificial Intelligence Laboratory, Cambridge, MA, 1986.

[2] R.J. Bobrow and J.S. Brown, "Systematic understanding: synthesis, analysis, and contingent knowledge in specialized understanding systems," in *Representation and Understanding*, R.J. Bobrow and A.M. Collins, Eds. New York: Academic Press, 1975, pp. 103–129.

[3] R.A. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. 2, no. 1, pp. 14–23, 1986.

[4] R.A. Brooks, "A hardware retargetable distributed layered architecture for mobile robot control," in *Proc. IEEE Int. Conf. on Robotics and Automation*, Raleigh, NC, 1987, pp. 106–110.

[5] R.A. Brooks, "Intelligence without representation," *Artificial Intelligence*, vol. 47, no. 1–3, pp. 139–159, 1991.

[6] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation," *Artificial Intelligence*, vol. 99, no. 1, pp. 21–71, 1998.

[7] L. Steels, "Cooperation between distributed agents through self-organization," in *Decentralized AI*, Y. Demazeau and J.P. Müller, Eds. Amsterdam: North-Holland, 1990, pp. 175–196.

[8] R.C. Arkin, *Behavior-Based Robotics*, MIT Press, Cambridge, MA, 1998.

[9] M.J. Mataric, "Behavior-based robotics as a tool for synthesis of artificial behavior and analysis of natural behavior," *Trends in Cognitive Sciences*, vol. 2, no. 3, pp. 82–87, 1998.

[10] L.P. Kaelbling, M.L. Littman, and A.W. Moore, "Reinforcement learning: A survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.