Title	ロボットの統合発達アーキテクチャに基づいた人ロボットインタラクションの個別化に関する研究
Author(s)	Pratama, Ferdian Adi
Citation	
Issue Date	2016-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/13518
Rights	
Description	Supervisor:丁　洛榮，情報科学研究科，博士

Enforcing Personalized Human-Robot Interaction through an Integrated Epigenetic Robot Architecture

Ferdian Adi Pratama

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Enforcing Personalized Human-Robot Interaction through an Integrated Epigenetic Robot Architecture

Ferdian Adi Pratama

Supervisor: Professor Nak Young Chong

School of Information Science
Japan Advanced Institute of Science and Technology

March 2016

Abstract

This research describes a robot architecture based on the epigenetic approach that is able to model robot behaviors using the robot past experience and contextual information. When two humans interact, an *interaction gap* may arise between them when they refer to the same object, concept or event in the real-world, but they associate it with a different meaning. However, as long as the interaction progresses, the gap can be reduced by continuous interaction and adaptation to form a sort of mutual understanding. In human-robot interaction processes, the interaction gap can be present and it is difficult to reduce, given the limited capabilities of current robot architectures in knowledge acquisition, revision, and adaptation. We posit that it is possible to enforce mutual understanding between a human and a robot providing the latter with the possibility of building a *personalized* experience as far as the interaction with the former is concerned, and we propose a conceptual design and implementation of Epigenetic Robot Intelligent System (ERIS), a robot architecture that is capable of acquiring and revising relevant knowledge during the interaction process. Experiments are aimed at demonstrating how different robots when exposed to different stimuli and interaction processes, are capable of conceptualizing different past experiences and *memories*, and ultimately engaging humans in contextualized interaction.

Keywords: Epigenetic architecture, developmental learning, memory-inspired architecture, long-term knowledge acquisition, context-based memory retrieval

Acknowledgements

Growing up, robotics has been purely my sole interest that drives me in the pursuit of my academic career, until recently I realized I have a hidden passion in psychology and behavioral science for quite a long time. The interdisciplinary of these subjects motivates and drives me further to conceive ideas for this thesis work, and I would like to recognize all the people around me that help cultivate my personal development for that matter.

I am blessed to have Prof. Nak Young Chong as my advisor, and I would like to express my sincere gratitude to him, especially for his continuous support, guidance, and faith toward this research. I am grateful that I was given freedom to choose my own research topic that piqued my interest, and he let me take charge of the research direction. Over the years, there was a nontrivial amount of skepticism aimed toward this research from many opposing sides, as Developmental Robotics is relatively a new field in the community. In spite of that, he believed in the future prospects and the potential of this research, and supported me no matter what. He encouraged me to spend my minor research period at the University of Genova, which turned out to be a milestone in my research work. He has been an amazing mentor and advisor throughout my academic life in JAIST.

I also would like to express my profound gratitude to assistant professor Fulvio Mastrogiovanni, for the constructive feedbacks, insights for the research, and the continuous support. When I first met him during his visit to our lab in JAIST, I was taken aback when he told me that he was interested in my research, while it took me a significant amount of efforts to successfully convince other related researchers. The moment I deliver my research presentation to him, I am still impressed how he managed to see right through the roadmap as well as the potentials of the work, despite only the conceptual framework and mathematical formulation were partially developed at that time. As a mentor and co-advisor, his attitude and approaches of dealing with challenges in academia inspires me to be a better researcher, and more importantly, a better person that inspires others.

During my minor research period at the University of Genova for three months, while working on the experiment with their Baxter robot, I get to know a lot of brilliant people, which shaped my perspective on the collaborative aspect and problem solving approaches in the robotics community in general. I would like to thank the post-doc members of the robotics lab of the Department of Informatics, Bioengineering, Robotics, and Systems Engineering (DIBRIS) in the University of Genova, as well as the Ph.D. students and the European Master on Advanced Robotics (EMARO) students for the fruitful discussions, feedbacks, and conversations. I am very grateful to the lab members of Chong Laboratory for the great experience during my entire graduate study.

Finally, a big thank you to my parents for their endless love, unwavering support, and encouragement at all times. This dissertation would not have been finished without their blessing. I wish to dedicate this work to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> Knowledge is in the end based on acknowledgement.
>
> Ludwig Wittgenstein

When two humans who do not know each other interact, either verbally or through gestures, it may be observed that misunderstandings originate because of their different opinions, personality, wishes, past experience, and ultimately culture. Even if they are referring apparently to the same object, concept or event, they may associate it with different properties (e.g., because of differences in perception), meaning (e.g., for the two have a different past experience) or implications (e.g, because the context in which they frame it differs). However, this is what makes interaction between humans so compelling and engaging. In such cases, we say that an **interaction gap** exists between the two humans, which can be mitigated through continuous learning, understanding and adaptation on both sides during the interaction process. This result is referred to as a **shared understanding**, i.e., the fact that although different, a mutual understanding of objects, concepts or events can reach at least a qualitative consensus (Glenberg, 1997; Gibbs Jr, 2005).

In a Cognitive Science perspective, this phenomenon may be linked to the *Symbol Grounding* problem raised by Harnad (1990), also in its version rephrased in the Developmental Robotics framework by Stoytchev (2009) as the **Principle of Subjectivity**. The same *symbol* in two *similar* developmental architectures of a robot or any other computational devices may be associated with different objects, concepts or events in the real-world. In such a case, the interaction gap is still present, and it appears whenever the two entities must exchange information and react accordingly.

In human-robot interaction (HRI) processes, the presence of interaction gaps is particularly deleterious, because it leads to stereotyped and unnatural patterns of interactions perceived by the human side. Eventually, this causes distress and a lack of interest in the interaction itself, which leads to an impossibility of accepting the robot one is interacting with: people tend to lose interest in the interaction after some time or after interacting with the same robot multiple times. This evidence is reported by a number of surveys related to long-term human-robot interaction in a variety of domains, such as health care, robot-assisted therapy, and education, as well as in different environments, such as work and home settings, and public spaces (Salter, Dautenhahn, & Bockhorst, 2004; Gockley et al., 2005; Leite, Martinho, Pereira, & Paiva, 2008; Fernaeus, H$r$akansson, Jacobsson, & Ljungblad, 2010; Leite, Martinho, & Paiva, 2013).

However, there is a fundamental difference between human-human and human-robot interaction dynamics. Between humans, the interaction gap can be usually reduced as long as the interaction progresses through a continuous learning and adaptation process: the two have to get acquainted and to trust each other. A similar process involving humans and robots has not been considered in detail yet.

1.1 Thesis Problem

In order to foster the debate around such issues, the work presented in this dissertation aims at designing a robot architecture based on the epigenetic paradigm, which considers the specific robot experience and the context framing the interaction with a human as prerequisites for human-robot interaction processes. To this aim, we argue that a robot should exhibit a kind of *individuality* originated from *artificial personality*, i.e., a purposive interaction with humans based on both a robot-specific experience and a given context.

In the past few years, various research efforts have been put forward to understand the role of robot behaviors and artificial emotions to form a sort of robot personality, as perceived by humans during the interaction process (Miwa, Umetsu, Takanishi, & Takanobu, 2001; Lee, Peng, Jin, & Yan, 2006; Chastagnol, Clavel, Courgeon, & Devillers, 2014; Tielman, Neerincx, Meyer, & Looije, 2014). Surprisingly enough, this issue has not been considered from a developmental perspective. With the aim of allowing both the bootstrap of a robot-specific experience and the use of contextual information when interacting with humans, the need arises for an open robot architecture based on the epigenetic paradigm that supports these two requirements.

On these premises, we identify the thesis problem from the challenging issues of the research work as follows:

> **Thesis Problem:**
> Interaction gap present during human-robot interaction is difficult to mitigate, in particular, when the robot's AI framework is not designed according to developmental paradigm.

1.2 Thesis Statement

To solve the thesis problem, we state our thesis statement as follows:

> **Thesis Statement:**
> We posit that it is possible to reduce the interaction gap by enforcing mutual understanding between a human and a robot, in particular, providing the latter with the possibility of building a *personalized* experience as far as the interaction with the former is concerned.

1.3 Thesis Approach

Based on the thesis statement, we propose a robot architecture called **ERIS (Epigenetic Robot Intelligent System)**, which is characterized by the following features:

- **Autonomous bootstrap and continuous adaptation of knowledge from the interaction with humans and the environment**, i.e., a robot-specific experience or memory. In principle, we expect that two identical robots exposed to different perceptions may end up in having different experiences. At the same time, if the two robots were exposed to the same stimuli, but as part of a different human-robot interaction process, their experiences should differ as well. This idea is closely related to the Principle of Subjectivity, as advocated by Stoytchev (2009).

- **Ability to conceptualize, consolidate and describe the previously acquired robot-specific experience** using contextual information provided by a human. If inquired about their experience, robots should provide a response of *their own* on the basis of past interactions with humans and the environment, as well as of the appropriate context framing each human request. Such a process is expected to adjust over time, thereby causing a continuous progression and adaptation of the robot-specific experience. As a matter of fact, this ability is aimed at enforcing the interaction (in terms of engagement) at the cognitive level in human-robot interaction processes.

On the one hand, these requirements assume the availability of models for creating robot-specific experiences, provocatively referred to as a **robot memory**. On the other hand, it is necessary to ground such memory models in an architecture integrating perception, representation and action. Currently, no integrated and *context-based* robot framework is available, which is aimed at a long-term human-robot interaction using a precise characterization of memory components and their roles, i.e., taking into account their interconnections. Furthermore, in spite of recent research activities on memory-inspired architectures (Nuxoll & Laird, 2004; Morse, de Greeff, Belpeame, & Cangelosi, 2010; Bellas, Faina, Varela, & Duro, 2010), which are based on individual memory components, no holistic approach has been devised to provide robot architectures with the necessary flexibility to efficiently deal with contextual information.

In order to validate ERIS, we have performed a number of human-robot and robot-environment interaction experiments involving the use of two identical robots that are exposed to different visual stimuli and interact with humans through simple sentences (in principle, also by means of simplified grammars), while consolidating specific events in their memory. Various objects are presented to the two robots on a table in front of them. The configuration of the objects is changed as time passes. In the meantime, a human interacts with the robots in order to revise their knowledge and to inquire them about their past experiences. The idea of an underlying revision process done by humans is aimed at mimicking a continuous cognitive interaction and knowledge assessment that usually occurs when two humans interact. In this case, the human counterpart just assigns labels to specific fragments of robot memory, be them sensory knowledge or predicative symbols. Then, the two robots are able to assess the *familiarity* of the detected objects on the basis of their *individual* past experience and the associated human-provided labels, both as part of perception (e.g., when detecting a previously seen object) and conceptualization (e.g., during the recollection of robot-specific experiences). Examples of possible inquiries that may be posed by humans include: *Have you been presented with a purple box before a green ball? Did you move the blue box to the right hand side?* or *How many objects do you recall that are related to a box?*

1.4 Thesis Outline

We begin this dissertation by providing a strong motivation of the research work and its impact to the research field, and then reviewing some of the efforts done to address the problem in developmental (epigenetic) robotics field from various perspective in Chapter 2. Then we discuss the fundamentals of human memory organization in Chapter 3, which is the main insight of the research work.

In Chapter 4, we introduce ERIS (Epigenetic Robot Intelligent System), that allows robot to progressively self-develop its knowledge through interaction. Extensive formalism of ERIS is provided, which interconnection between components and memory items are explicitly expressed and evident. Its implementation as a ROS stack is also elaborated, which emphasize the flexibility of integration with ROS-compliant general purpose robots. After explaining the formalism and implementation, we discuss the familiarity mechanism used (*tagging* process), including motivations, influence to robot personal experience, as well as the flow of sequence in Chapter 5. Four major memory processing phases are also discussed, to provide readers a better comprehension from the perspective of developmental psychology.

As the validation of our proposed work, we explore two different experiments. The first one, elaborated in Chapter 6, has an emphasis on one of the major features of ERIS, to allow robots to manifest personal experience and gather knowledge in a progressive fashion. Two phases of interaction are covered in the experiment, robot-environment and human-robot interaction. For the robot-environment interaction phase, the robot interacts with objects in a workspace, in particular, by displacing one object at a time to a different position within the workspace (hence resulting in a different workspace configuration), and observes the visual changes occurred in the workspace. This phase is then followed by the human-robot interaction phase, where a human may pose some questions, in the form of query-based simplified natural language, regarding the past events experienced by the robot.

The second experiment highlights one of the most significant principles of epigenetic robotics, *The Principle of Subjectivity*, addressed by Stoytchev (2009). The principle states that human interaction history and received stimuli affect the robot personal development and past experiences gained. Using ERIS, we validated that the Principle of Subjectivity through illustrative scenario with the corresponding procedures in Chapter 7. We finish in Chapter 8 with a summary of the dissertation along with contributions to the robotics community in general, especially to the field of epigenetic robotics. In addition, we provide some insights for the future research direction of the work.

Chapter 2

Literature Review

> Our problem, from the point of view of psychology and from the point of view of genetic epistemology, is to explain how the transition is made from a lower level of knowledge to a level that is judged to be higher.
>
> Jean Piaget

In order to reduce the interaction gap between humans and robots, we argue that it is necessary to provide robots with the capability of developing a sort of *personalized* experience of the interaction process, on the basis of a context defined by the interacting human. The proposed approach is loosely inspired by theories related to the organization of human memory, since memory is a fundamental aspect in human development and learning, and therefore a key factor in long-term interactions with other humans and the environment. In the following paragraphs, we discuss relevant cognitive architectures, as well as epigenetic architectures that exhibit autonomous development capabilities. Finally, we comment upon general concepts useful for a comprehensive cognitive theory of human-robot interaction processes.

2.1 Relevant Cognitive Architectures

Cognitive architectures are aimed at modeling a number of aspects of the human mind related to how information is stored and processed. The emphasis is on the interconnections among the components of the architecture, which are aimed at resembling specific functions of brain activity. The resulting models are typically *static*, meaning that they are neither learned nor adapted over time in a general fashion.

In a sense, the majority of the existing robot cognitive architectures are characterized by contradicting aims. On the one hand, their target is to mimic very general-purpose knowledge representation structures and processes associated with human cognition. On the other hand, in order to be effective, it is necessary to develop specific models and the definition of all the related assumptions. Models are then applied to specific (yet complex) problem domains, such as the *Tower of Hanoi* puzzle (Altmann & Trafton, 2002), relevant traits of human behavior during flying experiences (Byrne & Kirlik, 2005), as well as in tutoring (Lewis, Milson, & Anderson, 1987). It is noteworthy that these activities require a very modest interaction with other humans and the environment. If we consider Brooks' arguments in *Elephants don't play chess* (Brooks, 1990), we argue that the interaction with humans and the environment is fundamental for a wide range of robot cognitive activities, which fact is overlooked by available robot cognitive architectures.

Despite these similarities, the architectures described in the literature differ above all in their overall conceptual design. A few of them include SOAR (Laird, Newell, & Rosenbloom, 1987), Dual (Kokinov, 1994), ACT-R (Anderson, Matessa, & Lebiere, 1997) (and the predecessor ACT* (Anderson, 1983)), CLARION (Sun, 2006), OpenCog (Hart & Goertzel, 2008), and CHREST (Gobet & Lane, 2010), just to name a few. In this Section, we focus on those architectures that are most relevant to our work.

One of the most popular general-purpose cognitive architectures (not specifically aimed at robotics) is ACT-R, which has been proposed by Anderson (1993) after a number of previous releases, including ACT* (Anderson, 1983). The architecture of ACT-R is heavily based on features typically associated with human memory, such as priming and attentive processes. However, it has been used for more general cognitive tasks, such as *natural language processing* (Budiu & Anderson, 2004), and reasoning (Anderson, Reder, & Lebiere, 1996; Altmann & Trafton, 2002).

ACT-R advocates the difference between *declarative* memory (i.e., knowledge about facts) and *procedural* memory (i.e., a rule-based production system used to derive new knowledge from already available facts). Recent developments include the integration of connectionist elements within the main ACT-R architecture (Lebiere & Anderson, 2008).

The main limitation of ACT-R with respect to human-robot interaction processes is that the acquired knowledge cannot be easily adapted nor personalized, thereby being not useful at all in reducing the arising interaction gap. However, a notable theoretical result of ACT-R is the emphasis on *Rational Analysis*, namely the insight that human cognitive functions reflect a number of properties of the environment, i.e., the context. This is a key feature of our system as well.

OpenCog is a general-purpose framework aimed at implementing the notion of Artificial General Intelligence (AGI), i.e., to simulate relevant traits of human cognition and reasoning (Hart & Goertzel, 2008). OpenCog is still under development, but recently efforts have been made to integrate it with the robots by Hanson Robotics. However, differently from ACT-R, the architecture is organized in functional components, specifically dealing with (probabilistic) knowledge representation structures, attention, learning, natural language processing and emotional states.

In contrast with the approach taken in OpenCog, our architectural design is inspired by psychological studies of human memory. Furthermore, although OpenCog includes specific modules that may be used for human-robot interaction purposes, it does not consider knowledge acquisition, revision and adaptation, which is a core part of our approach and of the utmost importance to reduce the interaction gap between humans and robots.

2.2 On the Epigenetic Approach to Robot Architectures

The epigenetic (or developmental) approach to robot architectures foresees robots able to adapt their knowledge and to learn new facts, also interacting with humans and the environment. In the following paragraphs, we refer to general design concepts in epigenetic architectures that are relevant for our approach.

The notion of *Self-Aware Self-Effecting* (SASE) agent has been proposed by Weng (2004) to grasp what we have called interaction gap. In particular, Weng postulates the need for two different internal representation structures, namely the *world concept* and the *mind concept* structures. The former refers to objective knowledge, i.e., factual knowledge that is actually true in the real-world, whereas the latter is related to subjective knowledge, i.e., partial knowledge about facts and events that is accessible by the robot. Since, in order to ground symbols to actual perceptions, it is necessary to establish a mapping between the discrete, symbolic level and the continuous, numerical one, depending on their perception capabilities different robots may establish different mappings. Eventually, this fact may lead to different numerical and sub-symbolic representation structures for the same symbol. In a sensing perspective, the presence of different mappings is considered negative and it usually requires *ad hoc* calibration processes in order to uniform sensor responses (Denei, Mastrogiovanni, & Cannata, 2015; Youssefi, Denei, Mastrogiovanni, & Cannata, 2015) whereas, in a Cognitive Science perspective, the presence of different mappings corresponds to having different *meanings* for the same concept.

This argument is related to the so-called Symbol Grounding problem introduced by Harnad (1990), and later discussed by Stoytchev (2009) from an epigenetic perspective. While the basic version of the Symbol Grounding problem refers to the correspondence between symbols and represented entities in the real-world, in the case we consider, symbols originate as a consequence of the interaction process between robots and humans or the environment. As such, in our case the symbol grounding is performed by a human-assisted revision process occurring during the interaction with the robot. The result of the revision process is a semantic representation of the past robot experience.

Blank, Kumar, Meeden, & Marshall (2005) posit that self-exploratory behaviors be the central paradigm to allow robots to expand their repertoire of skills, without human supervision. Although the study of such emergent robot competencies is at the core of studies about intelligence, such approach may not be the most suitable solution for human-robot interaction processes, where a context associated with the interaction is usually available and defines at least part of the unexpressed semantics of the interaction itself.

Prince, Helder, & Hollich (2005) went further in this direction and introduce the notion of *Ongoing Emergence*, which introduces six fundamental requirements that must be exhibited by robots to be labeled as epigenetic. The six criteria are *continuous skill acquisition* using already available skills, features of the environment and the robot internal state; *integration of novel and existing skills*, in order to form a repertoire; *autonomous development and adaptation of skills* on the basis of current knowledge and goals; *bootstrapping of basic skills*; *stability of skills* over long periods of time; *reproducibility of skills* among robots undergoing a similar experience.

With respect to the criteria posited by Prince and colleagues, the one related to reproducibility plays a central role also in our architecture. As a matter of fact, we argue that the human-robot interaction process (also by means of the revision procedure) defines the specific robot experience that is exhibited in later stages.

Stoytchev (2009) identifies five principles that underlie any robot architecture based on the developmental paradigm. Among them, of particular interest for the architecture we propose is the Principle of Subjectivity, which originates from the Verification Principle. Since, according to the developmental approach, a robot is expected to bootstrap and verify its own knowledge, in doing so it can only verify it subjectively, which leads to a subjective, personalized, robot experience. It is noteworthy that this process imposes limitations on the amount of information that can be acquired by a robot. As argued by

Stoytchev, since a robot has to experience any relevant knowledge, time limits the amount of knowledge it can learn, unless *it relies on experiences provided by others*.

If we ground the learning process in a human-robot interaction framework, we achieve two important points raised by Stoytchev. In fact, the human-assisted revision process taking place during the interaction serves two purposes: on the one hand, as we discussed previously, it is expected to reduce the interaction gap because the robot experience would depend on human conceptualization; on the other hand, it allows us to reproduce an experiential teaching process that speeds up robot learning.

2.3 Missing Elements in Existing Epigenetic Architectures

As we discussed in the previous Section, the main principles of the epigenetic approach to robot architectures are open-ended learning, continuous knowledge acquisition, embodiment, and self-motivation. A few approaches in the Literature address one or more of these points. Although the systematic study of human memory traces back to 1960s (refer to the book by Squire & Kandel, 2000; Baddeley, Eysenck, & Anderson, 2014, for an extensive historical account and the references therein), only in the past few years a number of approaches have been presented, which aim at modeling different aspects of human cognition and reasoning, as well as at developing computational paradigms to encode them in robot cognitive architectures. In the following paragraphs, we limit our attention to literature explicitly taking memory components modeling into account, possibly grounded in a robot implementation. In the past few years, two approaches have been presented, which attempt at modeling architectural aspects of memory *as a whole*, namely the work by Morse, de Greeff, et al. (2010) and Bellas et al. (2010). Both the approaches put a great emphasis on memory components and their interconnections.

Bellas et al. (2010) consider a learning by evolution perspective using a *Multi-level Darwinist Brain* (MDB) proposed by Bellas & Duro (2004) to develop an evolutionary behavior-based robot architecture. The framework is based on an Artificial Neural Network (ANN) neuro-evolutionary approach. MDB aims at providing systems with life-long learning capabilities by adopting an evolutionary approach. The architecture potential is showcased using an AIBO robot, which learns a basic skill for catching a ball. Once the skill is learned the first time, the ball can be placed anywhere in robot sight. The framework allows for concurrent behavior execution and ANN-based continuous knowledge evolution. Unfortunately, the learning process can only be based on parameters specified in advance, such as the distance and the angle between the robot and the ball, as well as the ball's angular speed. Furthermore, a clear description of the advantages of applying evolutionary approaches to a robot cognitive architecture is not adequately motivated, and the proposed framework lacks much detail about the actual organization of the components, as well as their mutual relationships. At its core, MDB employs a multi-layer perceptron network, which limits self-developed knowledge acquisition. In fact, its configuration is adjusted in advanced and requires a manual tuning process to obtain optimal results. For this reason, we argue that MDB is not suitable for human-robot interaction processes, which require an autonomous adaptation of robot's knowledge.

The ERA architecture by Morse, de Greeff, et al. (2010) is a behavior-based robot architecture employing Kohonen's Self-Organizing Maps (Kohonen, 1998). ERA has been designed to exhibit a wide range of robot behaviors based on psychological traits. To assess the congruence between robot behaviors and such traits, so-called *modi* experiments are carried out using an iCub robot as the experimental platform (Parmiggiani et al., 2012; Smith & Samuelson, 2009). The modi experiment (also known as the *binding* experiment) is a procedure used in child language studies, where it demonstrates the role of embodiment in children early linguistic learning stages (Cangelosi & Schlesinger, 2014), and in particular their ability to correlate objects with words. In short, the procedure is as follows: first a child is presented with two objects in two distinct visual spots (e.g., on the left and the right hand side); after a short time, objects are removed from sight; having both visual spots free from cues, the child is introduced to the word "modi" at one of the two spots; after both objects are presented in a different location within the two available spots, the child is asked to find the modi; the majority of the children correctly select the correct object. In case of robots, the procedure is applied in a similar way, and the correlation is managed by a SOM structure.

Although thoroughly discussed from the Neuroscience perspective and the experimental procedure is later discussed by Morse, Belpaeme, Cangelosi, & Smith (2010), all the corresponding implementation details are left unclear. According to the authors, ERA is consistent with constructivist sensorimotor theories, and it also encompasses the *Echo State Network* framework (Yildiz, Jaeger, & Kiebel, 2012). The main drawback of ERA is the inability to capture temporal relationships and complex interaction patterns, which is significant in an epigenetic approach. This is due to the fact that both Self-Organizing Maps and Echo State Network configurations are defined in advance, which limits the development of new knowledge.

2.4 The Role of Context in Reducing the Interaction Gap

The role of *contextual information* in human and animal behavior is fundamental at various levels (Mehl & Conner, 2012). In humans, experimental evidence suggests that context-aware processes are represented mostly in the hippocampus (Smith & Mizumori, 2006). Given any situations, such processes allow for the execution of the most appropriate behavioral response or the retrieval of the most relevant memory item. Eventually, contextual information influences the way we create mental models of what we perceive and remember (Godden & Baddeley, 1975; Smith & Kosslyn, 2009).

To design robots able to proactively and sensibly understand their environment and to engage humans in long-term interaction processes, a similar concept has to be defined and integrated into its architecture. The proposed architecture assumes the onset of contextual information during the interaction process, and exploits such information to influence the retrieval of personalized robot experiences. When a human-robot interaction process occurs, we say that the process itself enables an interaction *context*, which defines implicit (social) rules and dynamics (Allen & Bekoff, 1999).

Since contextual information assumes the availability of mental models reflecting relevant traits of the interaction, and knowledge representation structures akin to memory, we emphasize in this Section relevant approaches aiming at computational models of human memory.

An analysis of the Literature shows that the focus is mainly on single memory components, such as the Working Memory (WM) (Phillips & Noelle, 2005), the Semantic Memory (SM) (Dodd, 2005; Dayoub, Duckett, Cielniak, et al., 2010), the Episodic Memory (EM) (Nuxoll & Laird, 2004; Dodd & Gutierrez, 2005; Kuppuswamy, Cho, & Kim, 2006; Jockel, Westhoff, & Zhang, 2007; Jockel, Weser, Westhoff, & Zhang, 2008; Nuxoll, 2007; Kasap & Magnenat-Thalmann, 2010; Stachowicz & Kruijff, 2012; Nuxoll & Laird, 2012; Tecuci & Porter, 2007) and the Procedural Memory (PM) (Salgado, Bellas, Caamano, Santos-Diez, & Duro, 2012).

Stachowicz & Kruijff (2012) provide a detailed explanation of both design requirements and formal concepts needed to characterize the episodic memory and its storage structure. However, the focus of their work is on the notion of *event* and its properties. Despite their claim of having designed a structure resembling the episodic memory, they do not take into account the notion of *context*, which is considered of the utmost importance discussed by Smith & Kosslyn (2009); Godden & Baddeley (1975) to frame memory consolidation.

When an attempt is made to design a more comprehensive memory-based robot architecture (Nuxoll & Laird, 2004; Morse, de Greeff, et al., 2010; Bellas et al., 2010), the goal is restricted to finding a solution to a very specific problem, rather than providing the robot with the capability to develop its own knowledge. Furthermore, neither the relationship between the different components is explicitly addressed, nor the mutual influence between components is usually considered. In any case, no clear use of the notion of context is provided.

2.5 Conceptual and Lexical Processes

As discussed by Martin, evidence suggests that the representation of real-world objects in the human brain is distributed Martin (2007). On the one hand, object properties (mostly related to their shape, e.g., exploiting visual and tactile information) are maintained within sensory and motor areas, as a result of the interaction with those objects. On the other hand, a categorical organization seems to be present as well, then encompassing more conceptual and even lexical layers. The conceptual representation corresponds to exemplar forms or basic categories, whereas the lexical representation refers to *labels* or references to the represented object.

To corroborate this insight, studies by Schacter and colleagues (Koutstaal et al., 2001; Simons, Koutstaal, Prince, Wagner, & Schacter, 2003) show that the repetition suppression phenomenon holds for both previously seen objects and, to a lesser extent, also for objects belonging to the same category but characterized by a different visual shape with respect to previously presented objects. In order to allow for a simple categorization mechanism, the outcome of the revision process is to provide robot knowledge with semantic labels, with the aim of exploiting labels to associate different (yet related) information.

2.6 Chapter Summary

This chapter draws a distinct borderline between the notion of *epigenetic* architectures (also known as developmental architectures) and *cognitive* architectures. This is considered to be significant, as the two architectures serves different purposes: cognitive architectures are aimed to simulate specific human cognition in a robotic system, while epigenetic architectures are aimed to provide the system the capability to develop intellectually in general. We also discussed profound conceptual and philosophical proposals from other researchers, and what is missing from the currently available epigenetic architectures. In principle, from the developmental psychology perspective, we consider the role of context to be significant in reducing the interaction gap, which is currently missing in the available epigenetic architectures. The fundamental philosophy of human development supported with these related findings motivates us further to enforce the mutual understanding between a human and a robot through a developmental system. As natural phenomena occurs during human development, the capability of progressive knowledge acquisition, revision and adaptation, are considered to be the main contribution and features of the proposed system.

Chapter 3

High-level Memory

> Without memory, there is no culture.
> Without memory, there would be no
> civilization, no society, no future.
>
> Elie Wiesel

In this chapter, we discuss the fundamentals of human memory organization, which is analogous to the conceptual and formulation proposed here, specifically designed for ERIS.

3.1 Overview

Study of human memory suggests that memory consists of three major components: Sensory Memory, Working Memory (WM - previously called Short-Term Memory), and Long-Term Memory (LTM). The interplay between these components determines whether stimuli received by five senses and either put them to be processed, stored for later used, or even ignored eventually. As depicted in Figure 3.1, LTM is categorized into explicit (declarative) memory and implicit (non-declarative) memory. Explicit memory stores information that can be recalled in a conscious fashion. This includes facts about the world and personal experiences (experienced past events relative to a particular time and space). On the other hand, implicit memory stores procedural-like and sort of abstract information that cannot be consciously recalled. There are quite a lot of variety information that can be stored in implicit memory, including procedural-like motor skills (e.g., how to ride a bicycle), perceptual priming (e.g., people tend to guess the word "prime" when a word-stem "_rime" is presented, if the word "prime" was displayed before the

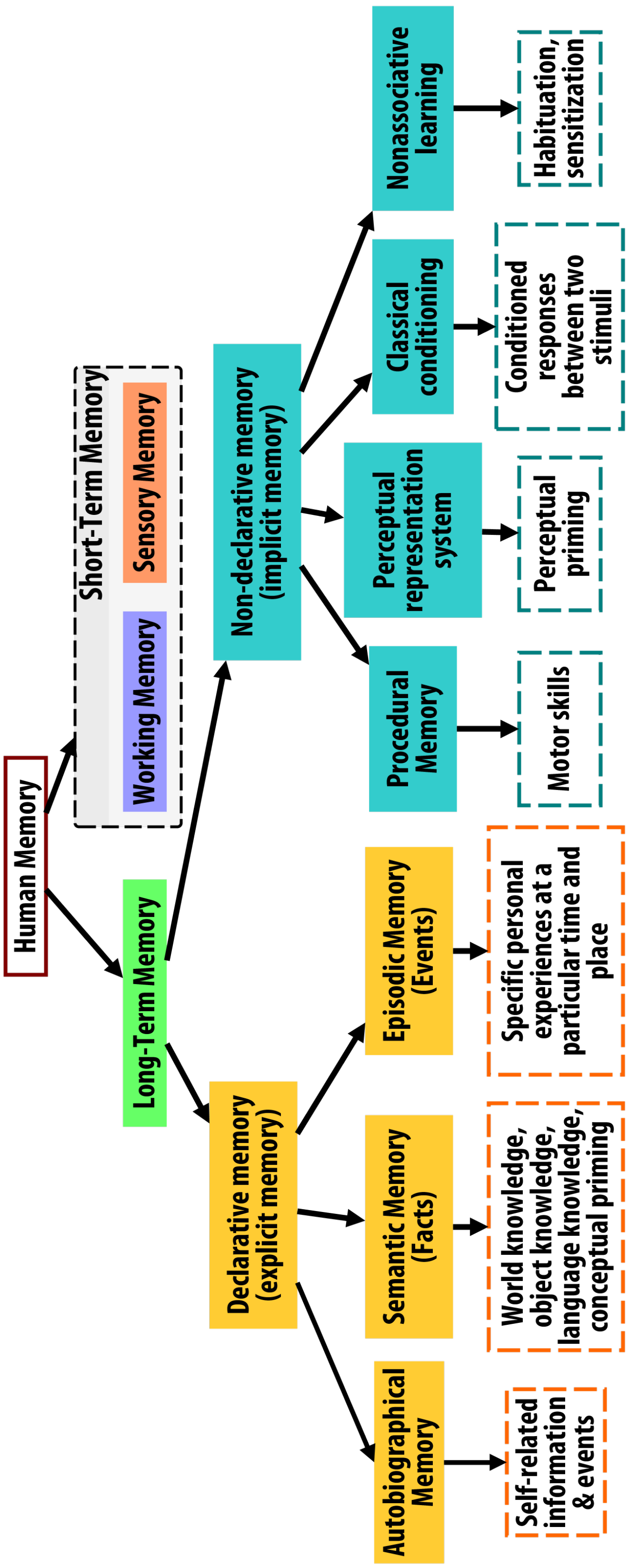

Human Memory
Short-Term Memory
Working Memory
Sensory Memory
Long-Term Memory
Declarative memory (explicit memory)
Non-declarative memory (implicit memory)
Autobiographical Memory
Semantic Memory (Facts)
Episodic Memory (Events)
Procedural Memory
Perceptual representation system
Classical conditioning
Nonassociative learning
Self-related information & events
World knowledge, object knowledge, language knowledge, conceptual priming
Specific personal experiences at a particular time and place
Motor skills
Perceptual priming
Conditioned responses between two stimuli
Habituation, sensitization

Figure 3.1: The hierarchy of human memory

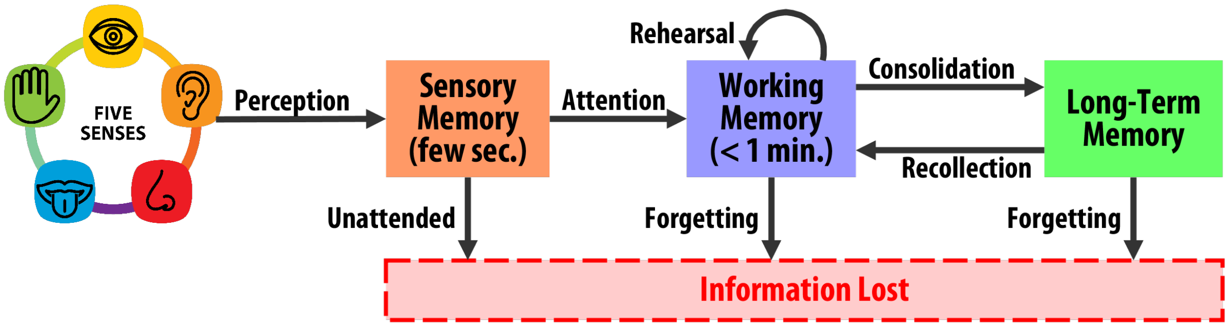


Figure 3.2: Human memory processing

word-stem), response between stimuli (e.g., skeletal muscle related stimuli), as well as habituation (e.g., decreased response due to a repeated stimulus) and sensitization (e.g., increased response due to a repeated stimulus).

An example of habituation is factory workers that already accustomed to the sound of machinery in the surroundings, and example of sensitization would be an annoyed person due to his neighbor keeps knocking on his wall for no reason. Habituation tends to be a stimulus-specific, and sensitization is rather a stimulus-general.

Figure 3.2 depicts the memory processing between sensory memory, WM, and LTM. The stimuli from five senses perceived through sensory memory. Here, stimuli is processed into raw information corresponding to each senses in less than a second. Then, if a stimulus is attended, it will go to the WM to be processed consciously. Unattended stimulus will result in information loss. In the WM, a component called *Central Executive* is believed to be managing all the processing details, including consolidation of memories, recollection, as well as rehearsal. The process in the WM occurs in a short time (less than a minute), a little bit longer than the sensory memory. Rehearsal process is a means to maintain the information to be processed, since unrehearsed memory leads to information loss. Rehearsal process can be either physically (e.g., verbally repeating the desired information to be remembered) or mentally occurs (e.g., keep thinking the same idea). A consolidated memory within LTM storage can be used for future use, until it is forgotten.

3.2 Sensory Memory

This kind of memory deals with initial process of storing information that is perceived through our five senses. It lasts for a very short period of time and being replaced con-

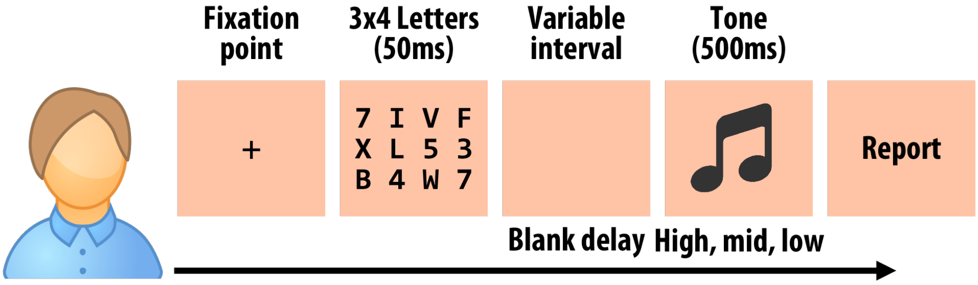


Figure 3.3: The iconic memory partial report experiment.

stantly as our senses work continuously. This short period varies among the different kinds of sensory memory. Due to the massive amount of information being processed, it is virtually impossible to consciously recognize all of this information. As sensory memory channels a lot of details within a short amount of time to our brain, there exists five different kinds of sensory memory corresponds to our five senses that are fed with these details of information very rapidly. When information that go through sensory memory are attended, they end up consciously processed by our brain in WM and consolidated into LTM. This leads us to the fact that, sensory memory allows us to decide whether an information is useful and make a quick reaction (e.g., physical reflexes) and judgement without having to wait for the information to be processed in WM.

There are three kinds of sensory memory that are main topics in recent study of the subject: Iconic Memory, Echoic Memory, and Haptic Memory; corresponds to visual, auditory, and touch senses. This is because, although all five senses influences the sensory memory, only visual, auditory, and haptic stimuli received more attention from researchers. The sensory memory related to the stimuli obtained by smell (olfactic memory) and taste (gustic memory) are rather enigmatic at the current research development of the topic.

3.2.1 Iconic Memory

Iconic memory deals with visual sensory information. An early experiment conducted by Sperling (1960) to see how many letters his subject could read during the brief flash of a tachistoscope. Seven years later, the term *iconic memory* coined by Ulric Neisser in his book (Neisser, 1967). During the early days, a tachistoscope was a device that displays an image for a specific amount of time. It was created to improve people's reading speed or enhance memory, and often used in psychological experiments related to visual stimuli. Although seven experiments were conducted, here we highlight the major two experiments. For extensive details of the rest of the experiments, readers are encouraged to refer to the publication (Sperling, 1960). In the first experiment (also called *whole report* condition), the tachistoscope was used to display various grid arrangement of letters (i.e., 1×3, 2×3, 3×3, 3×4, etc.), to his subjects for a $50ms$ of exposure. The result suggests that on average, the test subjects could read three to four letters in the 3×4 letters arrangement. The second experiment was the extension of the first one. It is also called *partial report* condition, because the subject needs only to report the requested row from the given letters in the grid arrangement. This experiment is designed to make sure the partial report is four letter or less, which lies within the subject immediate-memory span. This was done in a similar fashion with the first experiment only for 3×3 and 4×4 grid arrangement of letters, with the addition of tone introduced after the letters are displayed, depicted in Figure 3.3.

The variety of the tone depends on the number of rows in the letter grid arrangement, i.e., high ($2500hz$), medium ($650hz$), and low tones ($250hz$) for both arrangement, in particular, a more differentiated medium tones for the 4×4 arrangement). Overall, the subjects are able to successfully report three to four letters of the requested row. Sperling came to a conclusion that the subjects are able to capture a visual representation of the whole grid of letters in a fraction of a second, which then later still accessible to be recalled after the tune was heard. As the visual image fades in a fraction of a second, the legibility of the content decreases, which eventually decrease the recollection accuracy as well. Another experiment by Sperling (1963) is about process called *masking*, where the perception and/or storage of the stimulus is influenced by external factors. The external factors can be either occurred before the presentation (forward masking), or after the presentation (backward masking). In the experiment, he found out that the higher the brightness level during the interval seems to interfere with the memory trace.

3.2.2 Echoic Memory

Echoic memory, which also coined by Neisser (1967), corresponds to the auditory sensory information. After Sperling's model of iconic memory became popular, researchers started to find out the auditory counterpart of sensory memory. A little bit different from iconic memory where it scans stimuli in a continuous manner, echoic memory does not (Carlson, Heth, Miller, Donahoe, & Martin, 2009). Auditory stimuli are received once before being processed and consciously understood. As a matter of fact, once heard, auditory stimulus resonates in our mind and replayed until it fades, which lasts up to 4 seconds, based on the finding from Darwin, Turvey, & Crowder (1972). Another result from a variation of the experiment, conducted by Glucksberg & Jr. (1970), suggests that the mental resonance of the stimulus lasts up to 20 seconds without interference. This makes echoic memory has a slightly longer duration than iconic memory.

An experiment about remembering patterns of telephone number is conducted by Murdock (1967). The result suggests that using visual presentation of the numbers systematically increase the likelihood of error in remembering the numbers from the beginning to the end of the sequence; while when aurally presented, it is more likely to be correct at the last one or two numbers compared to the rest of the sequence. A more recent research by Alain, Woods, & Knight (1998) suggests that damage to the frontal lobe, parietal lobe, or hippocampus, may have negative impact on echoic memory (shorter span or slower reaction time).

3.2.3 Haptic Memory

Haptic memory correspond to the sensory information acquired by touch. In general, haptic is classified as tactile and kinesthetic (Lederman & Klatzky, 2009). On the one hand, tactile refers to the perception on the surface of the skin, such as touch, pressure, texture, and vibrations. On the other hand, kinesthetic refers to the perception related to muscle, tendons, and joints. For instance, while holding a coffee-mug in your hand, you can estimate the physical properties of the mug (i.e., size and weight) and how the mug is held relative to your body.

During the early days, the study by Bliss, Crane, Mansfield, & Townsend (1966), provides insights about brief tactile stimuli applied to hand, which the procedures are inspired from the iconic memory experiment by Sperling (1960) (whole and partial report). They found out that the performance during the partial report was significantly improved. The study was then later backed up with additional supporting evidence by Gilson & Baddeley (1969). Gordon, Westling, Cole, & Johansson (1993) studied haptic memory in relation to interaction with environment by the means of assessing and adjusting gripping force. Another recent study by Shih, Dubrowski, & Carnahan (2009) suggests that the duration of haptic memory is less than 2 seconds, which makes the duration and decay similar to iconic memory.

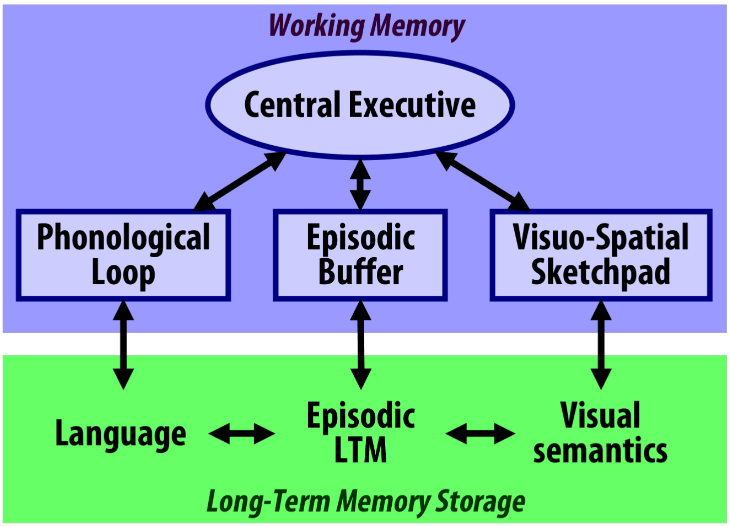


Figure 3.4: Baddeley's updated Working Memory model

3.3 Working Memory

The term of *Working Memory* and *Short-Term Memory* seems often to be used interchangeably. However, it is advisable to know the difference between the two. The term *Short-Term Memory* is neutrally used for a temporary storage involving retention of small information for a short period of time. On the other hand, *Working Memory* can be think as a system of a mental workspace that actively maintain and also process information temporarily, such as reasoning, learning and comprehension. Up to now, the WM model by Baddeley (2000), which is an updated model based on a previous design by Baddeley & Hitch (1974), is influential and widely used in all studies in neuropsychology about human memory. Figure 3.4 depicts Baddeley's updated WM model (Baddeley, 2000).

3.3.1 Memory Span and Chunking

One of the metrics to measure the performance of WM is the **memory span**. Fundamentally, to determine the memory span, we should know two things:

1. remembering what the items are
2. remembering the order of the presentation.

The term *item* here is rather vaguely and subjectively interpreted, and this is closely related with a term called *chunking*. Chunking is the ability to cluster into something more useful. The usefulness itself is something that is relative and different for one person to another. Tulving (1962) addressed this chunking process as *subjective organization*. For example, try to remember this: UESFSLSUNE. It is rather difficult to remember the letters and their order of presentation. Using the same set of letters, try again with this sequence of letters: USEFULNESS. This time, it is very easy to remember them. This is because our brain tries to cluster the letters into something more useful called a *chunk* (in this case, a word-like groupings of letters).

In the first case, since the letters are presented in a rather jumbled order, nothing useful can be extracted from the sequences of letters. Therefore, our brain tries to remember the exact letters one by one, as well as the order of the presentation. Eventually, since they cannot be clustered further into something more useful that retains the presentation order, ten chunks should be remembered, with a chunk representing a letter. In the second case, the sequence of letters is able to be grouped into a word, since our brain recognize the word “usefulness”. Therefore, instead of remembering ten items, our brain remembers only one chunk representing the word “usefulness”, which is much easier. The chunk itself already representing all the letters and the order of the presentation, because it linked back to our general knowledge stored in the LTM about the word “usefulness”. According to Miller (1956), the number of chunks that needs to be remembered influenced the memory capacity, instead of the number of items. This also implies that LTM can influence WM.

3.3.2 The Phonological Loop

According to Baddeley & Hitch (1974), the phonological loop consists of two subcomponents:

1. short-term store

2. articulatory rehearsal processing module

The short-term store duration lasts for a few seconds, and stores only memory traces of audio information. The purpose of articulatory rehearsal processing module is to refresh these memory traces. Aside from the auditory stimuli, visual stimuli can be manually transformed into phonological code through articulation, and therefore processed in the phonological loop instead of visuospatial sketchpad. There is also one popular finding about phonological loop, which is the psychological similarity effect, which is related to chunking process. Conrad & Hull (1964) demonstrated that lists of words that sound similar are more difficult to remember than words that sound different. This effect does not apply for similarity in meanings of the words. For instance, as from an experiment of Baddeley (1966b), recalling pit,day,cow,pen,top or big,wide,large,high,tall is easier compared to mad,can,man,mat,cap.

Baddeley (1966a) explained that under this circumstances, the semantic meaning plays an important role to determine the memory span. The second popular finding is by Baddeley, Thomson, & Buchanan (1975) about articulatory suppression, which states that something irrelevant that is repeatedly said (e.g., the word "the") may block the articulatory rehearsal process, which makes all the memory traces in the phonological loop to decay.

3.3.3 Visuospatial Sketchpad

Visuospatial sketchpad is the WM component based on the model of Baddeley & Hitch (1974), responsible for facilitating the processing of visual stimuli, including image manipulation, visualization, and visual recollection. Visual imagination in our mind is related with the details of the memory (i.e., about what can we see) and spatial relationship (i.e., where they are located within the virtual environment). This component is influenced by the way our eyes movement to "scan" the scene in an underlying discrete manner, instead of a smooth, continuous flow, to recognize different parts of the scene. This means, although we "seem" to continuously scan the scene, our brain divide the movements into a series of brief eye movements. A study by Logie (1995), the visuospatial sketchpad is composed of two subcomponents:

1. visual cache (stores form and color of detected objects)

2. inner scribe (spatial relationship and movement information of detected objects)

3.3.4 Central Executive

As the only active components of the WM, central executive responsible in supervising all the process occurs within all other components. Norman & Shallice (1986) proposed that central executive has two modes of control scheme:

1. based on habit (i.e., autonomous and semi-autonomous control)

2. supervisory attentional system.

Driving car routinely everyday from your home to your office is a good example of a semi-automatic control of central executive. In this situation, the knowledge of “driving a car” is not something new, the procedures are already learned beforehand, and the experience faced during the process is less likely different than most of the time. These facts help us by requires less conscious control. Assuming there is a sudden, novel situation occurs during the driving process (e.g., a roadblock that force you to find an alternative route), the situation will be handled by the supervisory attentional system. The supervisory attentional system responsible for intervening the novel situation out of the habit, and finding strategies to seek these alternative solutions.

3.3.5 Episodic Buffer

Episodic buffer is the new component as part of the Baddeley (2000) model, which purpose is to explain the interconnection between the WM and LTM. As agreed by Baddeley & Andrade (2000), information processed in the WM is not solely depends on individual components, and they agreed that there should be a component that links all the memory components. The Baddeley & Hitch (1974) model is not able to explain that fact. A motivation for the addition of the component is due to the evidence obtained from patients with amnesia based on the study of Baddeley & Wilson (2002), which the patients were not able to to encode new memories to the LTM, but have a good memory recalling capabilities from STM.

The episodic buffer is assumed to be a multidimensional components that able to hold four chunks of information, episodes (or chunks) based on different dimension (either visual, verbal, semantic originated from perception, WM, and LTM) (Baddeley et al., 2014), and binds the information if necessary. Baddeley (2012) updated the model for the second time. The changes is that episodic buffer have a direct access to the central executive and the other two passive subcomponents, due to the evidence based on the study of Luck & Vogel (1997); Vogel, Woodman, & Luck (2001); Daneman & Carpenter (1980); Kintsch & Van Dijk (1977) suggesting that information in visuospatial sketchpad and phonological loop are accessible through the episodic buffer directly.

3.4 Long-Term Memory

Human brains are integrated with a passive storage called the *Long-Term Memory* (LTM), to store experienced past events that we can recall in the future. Long-Term Memory are not constant, and human often revise their memory as they gained experiences, i.e., merging with another memory or modifying the memory contents. For the majority of the human population, the memories stored in the long-term memory do not last forever. The exception applies to a substantially small portion of human population, who claimed to have *eidetic memory*, or also known as *photographic memory*. A relatively new findings by Hadziselimovic et al. (2014) suggest that

1. human brains store only information that considered to be important, and
2. less important memories are deliberately forgotten by our brain (in unconscious manner) in order to remain efficient and reduce mental burden

The findings implies that human brains are *actively* filtering those unnecessary information that we receive everyday, by forgetting them.

Long-term memory storage has different forms to store different kinds of memories. As shown in Figure 3.1, long-term memory is distinguished as explicit and explicit memory. On the one hand, explicit memory, also known as *declarative memory*, corresponds to all the memory that can be retrieved consciously. Two major divisions of explicit memory are Semantic Memory and Episodic Memory. Semantic memory stores mainly facts about the world, and episodic memory stores specific events in a particular time. On the other hand, implicit memory, also known as *non-declarative memory*, corresponds to all the memory that does not necessarily require conscious thought. The example for this kind of memory is the things that you learn by rote, e.g., related to muscle memory. Implicit memory consists of several divisions, such as Procedural Memory, Perceptual Representation System, Classical Conditioning, and non-associative learning.

3.4.1 Semantic Memory

Semantic memory provides us a specialized container to store memories about facts and general knowledge of the world. For instance, we know that one minute is equal to 60 seconds, and a dog barks. Those memories are categorized to be independent with respect to any particular event. A clear definition about semantic memory is provided by Binder & Desai (2011), which states that “it is an individual’s store of knowledge about the world. The content of semantic memory is abstracted from actual experience and is therefore said to be conceptual, that is, generalized and without reference to any specific experience”. A study by Wheeler, Stuss, & Tulving (1997) suggests that recollecting details of information from the episodic memory is related to conscious recollection process. On the contrary,

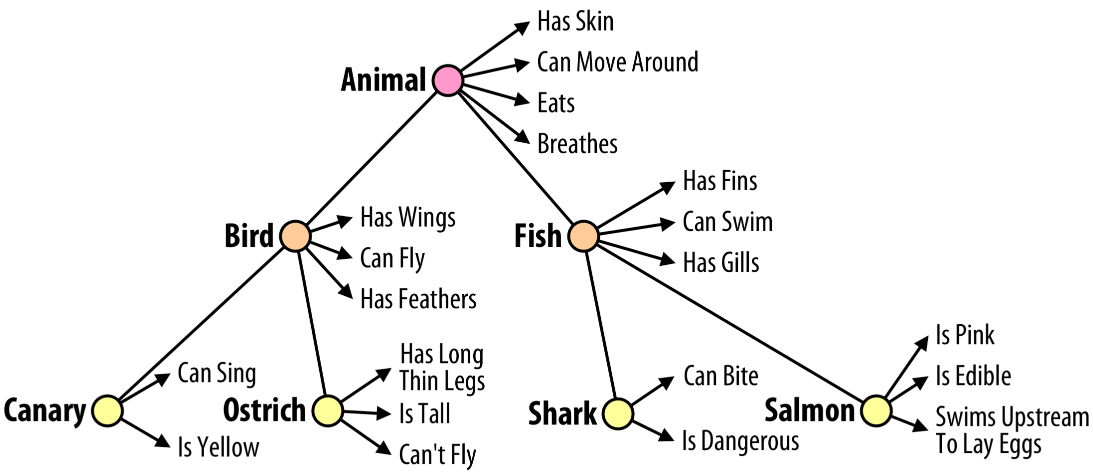


Figure 3.5: An example of hierarchical Semantic Memory model by Collins & Quillian (1969)

the semantic memory recollection can be performed in an unconscious manner. This fact is supported by additional evidence by Kan, Alexander, & Verfaellie (2009); Irish et al. (2011), which suggests that during episodic memory recollection, amnesic patients have more severe problems, and dementia patients have less severe problems; however, it is the opposite for both patients during the semantic memory recollection. The result of an experiment by Loftus & Suppes (1972) suggests that, for human, recalling words belong to a certain category using the first letter as the cue is faster than the last letter (e.g., recalling a fruit with the first letter *T* is faster than recalling with the last letter *t*). For computer, however, it is possible to create a filtering program that searches the desired words with either cues equally fast. The category fruit was used as a filter to limit the search domain, which speeds up the filtering process.

Collins & Quillian (1969) proposed a hierarchical model of semantic memory. The model revolves around the notion of *concept*, where one represented as a *node* and has one or multiple *properties* associated to it. The properties can be either general or specific. Each concept also hierarchically structured through association with other nodes, which example depicted in Figure 3.5. *Canary* is a concept node, and *is yellow* is one of the associated properties. The higher the node level, the more general the concept node is; and the lower the node level, the concept node becomes more specific. The statement such as *shark has skin* was recalled slower than *shark has fins* due to the gap between the hierarchy. Later, Collins & Loftus (1975) proposed a spreading activation model of semantic memory, because the hierarchy organization was considered to be inflexible. The spreading activation model was based on the notions of semantic relatedness and

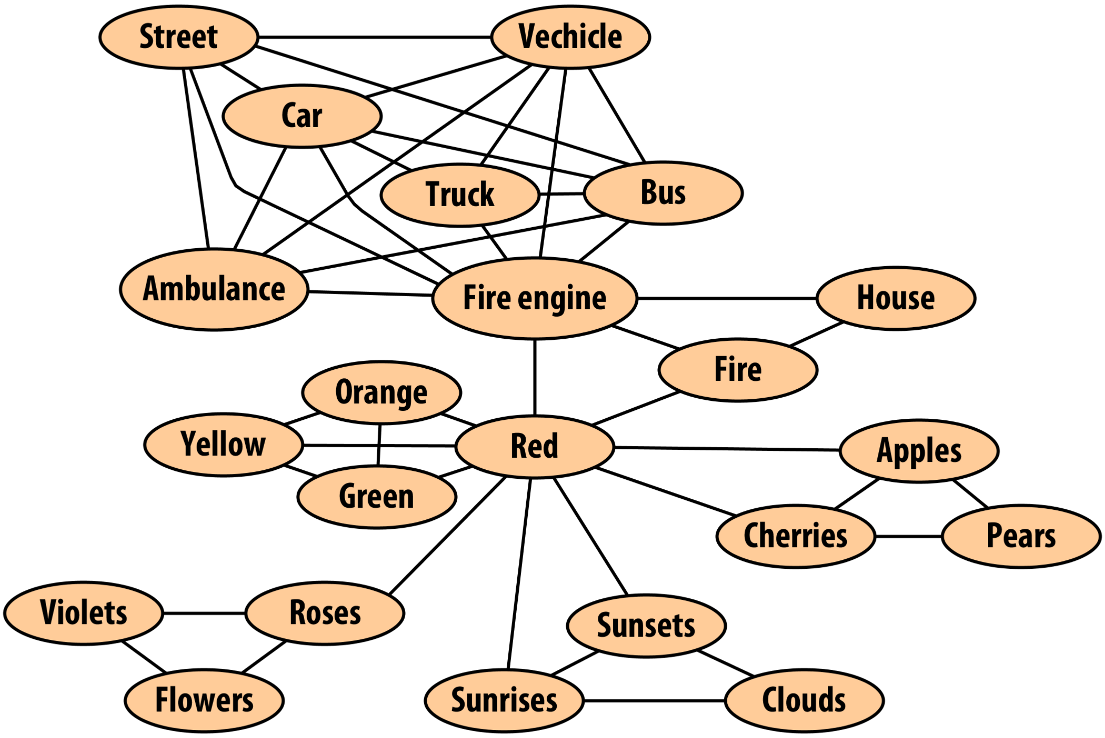


Figure 3.6: An example of spreading activation Semantic Memory model by Collins & Loftus (1975)

semantic distance, which is indicated by the distance between nodes. Both notions are measured by asking people to build the statistics. Shown in Figure 3.6, *red* is more related to *fire* rather than *sunrises*.

The spreading activation model is more flexible and successfully explain various phenomenon based on multiple findings. Despite the flexibility, the model yields less accurate prediction. Furthermore, the model oversimplified the representation of a *concept* into a single *node* with no associated individual properties, and with the assumption that a *concept* represent a single and fixed representation of an idea. An experiment by Chaigneau, Barsalou, & Zamani (2009) provides us some evidence along with insights about object categorization in two conditions: isolated and given a situational information, or in other words, contextual information. The result is contextual information improve the recollection accuracy.

3.4.2 Episodic Memory

When one starts to recall a specific event occurred in the past, the memories correspond to it are called Episodic Memory. Tulving (2002) defines it as a way to mentally time travel to

Table 3.1: The distinction between episodic-semantic memory according to Tulving (1972)

	Episodic	Semantic
Type of information represented	Specific events, objects, people	General knowledge or facts about the world
Type of organization in memory	Chronological (by time) or spatial (by place)	In schemas or in categories
Source of information	Personal experience	Abstraction from repeated experience or generalizations learned from others
Focus	Subjective reality: the self	Objective reality: the world

re-experience past events. Episodic memory considered not only to '*relive the past*', but as a platform to mentally simulate previously non-existent events, or to travel forward and plan events for future execution/occurrence. As semantic memory and episodic memory are considered a distinct memory storage, Tulving (1972) distinguish the difference between them based on several categories, listed in Table 3.1.

Baddeley et al. (2014) states that the accumulated events in the episodic memory might accumulate into the basic form of semantic memory. The proposal of Tulving (1972) and findings from Conway, Cohen, & Stanhope (1992) suggest that information within the episodic memory are recalled as *episodes*, and the recollection of episodes are somehow interconnected with the semantic memory. Closely related with the semantic memory definition previously by Binder & Desai (2011), they also provide one for episodic memory, which states that "it is a memory for specific experiences, although the content of episodic memory depends heavily on retrieval of conceptual knowledge. Remembering, for example, that one had coffee and eggs for breakfast requires retrieval of concepts of coffee, eggs and breakfast. Episodic memory might be more properly seen as a particular kind of knowledge manipulation that creates spatial-temporal configurations of object and event concepts."

Bartlett (1932) presented his students with North American Indian Folk Tales and asked them to recalled the story. He found that the recalled story was always shorter, more coherent and rather told from the third person/individual perspective rather than the original story. This indicates that instead of completely preserve the contents of the story and remember them as they originally told, all the listeners were clearly trying to find essential meaning of the story interpreted from their viewpoint. Categorized also as explicit memory, there is interconnection between episodic memory and semantic mem-

Table 3.2: Ten characteristics of episodic memory according to Conway (2005)

1	Retain summary records of sensory-perceptual-conceptual-affective processing derived from working memory.
2	Retain patterns of activation/inhibition over long periods.
3	They are predominately represented in the form of (visual) images.
4	Represent short time slices, determined by changes in goal-processing.
5	Represented roughly in their order of occurrence.
6	They are only retained in a durable form if they become linked to conceptual autobiographical knowledge. Otherwise they are rapidly forgotten.
7	Their main function is to provide a short-term record of progress in current goal processing.
8	They are recollectively experienced when accessed.
9	When included as part of an autobiographical memory construction they provide specificity.
10	Neuroanatomically they may be represented in brain regions separate from other (conceptual) autobiographical knowledge networks.

ory. This is best illustrated by an example as follows: imagine yesterday you ordered a spaghetti and a cup of coffee at a restaurant. Now, the process of remembering that event corresponds to the episodic memory recollection. In addition, remembering the episodic memory automatically interconnected with the semantic memory of the detailed information of that event. In this case, when you remember the event, you also remember about the coffee and spaghetti that served during that event. This interconnection reminds you the taste of coffee in general or the visual of a spaghetti typically served from your own experience, which makes it possible for you to compare the taste and determine which one is more delicious, and so on.

Conway (2005) defined ten characteristics of episodic memory, listed in Table 3.2. The characteristic number 8, recollective experience, is also known as *autonoetic consciousness*, which refers to the *mental time travel* discussed by Tulving (2002). It also means that the memories is often experienced in the form of imagery or other sensory-perceptual details as the original event during the recollection process. Episodic memory is somehow related with Autobiographical memory, as we will discuss in the next subsection.

Another influential finding is called the *dual-coding theory*, which is closely related to mental imagery. The theory states that human brain uses two representations to represent information stored as memory: visual and verbal information. Those two information are processed differently with our mind (Sternberg, 2003). For instance, the lexical stimulus “dog” can be encoded and recalled visually as a dog image, or verbally as the word *dog*.

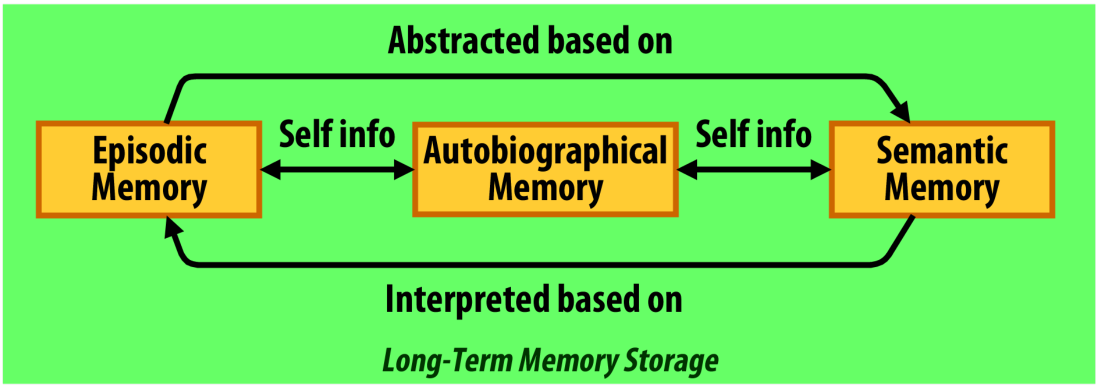


Figure 3.7: Interactive and interdependent relationship between episodic-semantic memory

This process tends to be more difficult as the stimulus becomes more abstract, such as the word *life, knowledge,* or *honor*.

3.4.3 Autobiographical Memory

As we have learned the distinction between the episodic memory and semantic memory in Table 3.1 due to their different purposes, studies by Dritschel, Williams, Baddeley, & Nimmo-Smith (1992) and Conway & Holmes (2005) suggest that both episodic memory and semantic memory are in an interactive and interdependent relationship, instead of a completely separate structures. Such relationship is clarified due to the fact that semantic knowledge is derived from personal experiences by abstraction and generalization process, and the recollection or interpretation of episodic memory is based on the general semantic knowledge in the form of schemas or categories, depicted in Figure 3.7. Researchers then came up with the term Autobiographical memories, which is a declarative memory originates from the relationship between the episodic-semantic memory in Figure 3.7 and self-related information.

Williams, Conway, & Cohen (2008) identified several dimensions of autobiographical memory:

1. Autobiographical memories may sometimes consist of biographical facts, for example, I may remember the fact I was born in Liverpool without having any actual memory of having lived there. Tulving defined this type of factual information as *noetic memory*, in contrast to *autonoetic memory* which is experiential. An example of an autonoetic-autobiographical memory is that I can relive the past experience with sensory imagery and emotions when I recall that I went to school in Wales.

2. Brewer (1986) argues that consolidated memories are, to a certain degree, either copies or reconstruction of the original event. This is due to some personal memories are vivid and remembered in a great detail, and some are not so accurate. Also, instead of being raw experiences, they incorporate the interpretation of the subject regarding the event. He also argues that it is plausible that noetic memories is more likely to be reconstructed.

3. Autobiographical memories may be *specific* or *generic*, for example, I may remember eating lunch at a particular restaurant on a particular occasion, or I may have a generic memory of family dinners. Neisser (1986) noted that a personal memory may be one that is a representative of a series of similar events, and he termed this type of blended memory as *repisodic*.

4. Autobiographical memories may be represented from an *observer* perspective or from a *field* perspective. A finding from Nigro & Neisser (1983) suggests that when people examined their own memories some were remembered from the original viewpoint of the subject (the field perspective), but a larger number of memories seemed like viewing the event from the outside, as if recalling them from an external observer viewpoint. As stated in the item 2 about the copy and reconstructed memories, Nigro and Neisser are convinced that memory from an external observer perspective must have to be reconstructed, and cannot be copies of the original event. They also reported that recent memories were more likely to be copy-type memories re-experienced from the original viewpoint, but older memories were more likely to be reconstructed ones seen from the observer's perspective. Robinson & Swanson (1993) replicated this finding and noted that field memories were more vivid, as it is considered a copy from the original event instead of the reconstruction ones. They asked students to recall personal memories from different period of their lives, to report on the perspective of each memory recollection. They also found that changing perspective from a field to an observer perspective or to recall it from other perspective is possible, but has diminishing effect, and it was harder to switch perspective if the memory was old and not very vivid. These evidence are consistent with the reconstructive theories as discussed in item 2.

Based on various evidence and studies, Williams et al. (2008) defined three functions of autobiographical memory, which has a directive, social, and self functions.

- ***Directive***
  As addressed by Baddeley (1987), the directive function of autobiographical mem-

ory involves using memories of past events to guide and shape current and future behavior, as an aid to problem solving, and as a tool for predicting future behavior. This is due to the fact that general knowledge abstracted from past experience may not always be relevant, and to solve the problem it may be useful to search back autobiographical memory to find a specific experience where similar problem was encountered. Pillemer (1998, 2003) elaborates the significance of this directive function and provides example of how autobiographical memory directed behavior on a large scale through a discussion on how terrorist attack on the World Trade Center on September 11, 2001 changed the behavior of Americans. Weeks after the tragedy, Pillemer (2003) found out that Americans chose not to travel by air and avoided public places for fear of their personal safety. He emphasized that although *facts* of the tragedy may affects the behavior, the personal autobiographical memories of seeing the horrific images of the collapsed Twin Towers on television has a direct impact on intensifying the reaction of the behavior.

- ***Social***
  Neisser (1988) consider this social function of autobiographical memory to be *the most fundamental function of memory*. As a social being, people share memories as a conversational material and exchange personal narratives to reduce interaction gap, which makes autobiographical memories a perfect means to facilitate social interaction. As of the studies by Fivush, Haden, & Reese (1996) and Bluck (2003) suggest, they discuss the self-disclosure of autobiographical memories in two following conditions; (1) sharing autobiographical memories with someone who was not present at the original event is considered to be a means of increasing intimacy in multiple ways, including pooling experiences and exchanging sympathy, and a way of “placing ourselves” in a given situation, culture, and context. (2) sharing autobiographical memories with someone who was present at the original event is considered to be a means of social bonding and to increase intimacy between the two sides of individual. Other numerous researches in this area including the study by Robinson & Swanson (1990) which suggests that social relationship may suffer when episodic remembering of autobiographical memory is impaired; and the study by Neisser (1988) and Nelson (2003) addressing the relationship between the social function of autobiographical memory with the potential evolutionary adaptivity.

- ***Self***
  Conway (2005) considers that autobiographical memory has a significant role to

the notion of “*self*” of an individual, which is: events that are remembered are of personal significance and are the database from which the self is constructed. It is also viewed as an essential element that build one’s personal identity. The notion of *Autobiographical knowledge* is proposed by Conway (2005) to constrain what the self is, has been, and can be in the future. There was also a study case by Scogin, Welsh, Hanson, Stump, & Coates (2005) where patients experienced memory loss through trauma or disease were not able to recall their own personal history, which makes them lose their sense of *self*. Woods, Spector, Jones, Orrell, & Davies (2005) treated patients with dementia using reminiscence therapy, which found to be improve cognition, mood, and general behavior functioning.

3.4.4 Procedural Memory

Procedural memory is a passive storage as a part of the long-term memory, responsible for storing instructional memory of how to do things, such as motor skills related memory. Different from the explicit memory (episodic memory and semantic memory), the memories within the procedural memory are accessible without the need of conscious thought. Some examples of procedural memories are: riding a bicycle, playing a piano, playing hockey, building a machine, fixing a radio. Despite the differences from explicit memory, procedural memory is interconnected to both episodic memory and semantic memory. One of the characteristics of procedural memory is that it is more difficult to verbalized for most people and verbal interpretation for each person is most likely different for most of the time. For instance, describing a particular event happened in the past is easier than describing how to play a piano, and the verbal explanation of building the same machine is less likely to be the same for two different people.

In human brain, it is believed that procedural memories are formed through reinforcing the same stimuli multiple times, which is something that we learn by rote, instead of through an explicit, conscious memory consolidation. Early studies by Milner (1962) reported from her experiment with severely amnesic patient H.M., that he was able to learn hand-eye coordination from skills because he previously learned the skills, without having any explicit memory. Since procedural memory can be recalled without conscious thought, we can say that in some sense, procedural memory defines our behavior and determines how we react to something. This also allows us to perform multitasking, for instance, ironing a shirt and talking to someone on the phone at the same time.

3.4.5 Perceptual Representation System

A common example of perceptual representation system is *priming*. It is a technique used in psychology to train a person's memory. Similar to procedural memory, priming occurs below our conscious level of perception and still persist for a long period of time even after single experience (Squire & Kandel, 2000). Priming can be used both in positive and negative way. A positive priming can be performed by exposing images or words to an individual. This is done to trigger the stimuli and the associated memories for the future memory recollection. This is influenced by spreading activation model of semantic memory.

A negative priming, on the contrary, slows down the memory process. This can be done by exposing stimuli which are not related to the desired stimulus to be remembered later in the future, and when the brain tries to remember, the stimuli interferes with the memory recollection process. In other words, priming heavily influences the subject's perception by either improving recent encounter stimulus for future memory recollection or impairing desired stimulus by exposing stimuli non-associated with the desired stimulus. In human brain, perceptual priming occurs in the posterior cortex.

3.4.6 Classical Conditioning

A popular example of classical conditioning is muscle memory. Muscle memory is the one responsible to the phrase "Practice makes Perfect" in our daily lives. It helps us learning physical activities by repeating a particular section of that activity. According to Bruusgaard, Liestøl, Ekmark, Kollstad, & Gundersen (2003), muscles cells are commonly large, especially in the vertebrate body, possess multiple nuclei, and constitute one of the few syncytia in the mammalian body. In other words, muscles cells are one of the few multi-nuclei cells in our bodies. A study by Bruusgaard, Johansen, Egner, Rana, & Gundersen (2010) shows that by overloading your muscles through strength training, new nuclei are added to the muscles, before any major increase in size occurs during the overload. In fact, according to an evidence by Gundersen (2011), muscles size regulation is determined by the number of nuclei in the muscle tissues. The internal mechanism of muscle memory itself does not distinguish whether what you learn is good or bad, therefore it is beneficial for one to take some time when training the muscle memory, to make sure that the procedure learned through repetition is the "good" one.

3.4.7 Non-associative Learning

Habituation and sensitization are the forms of a learning related to Long-Term Memory, which closely related to the response of a repeated stimulus. The most widely acknowledge contributions came from the study by Groves & Thompson (1970) and Thompson & Spencer (1966). As mentioned in Section 3, habituation is a decreased response due to a repeated stimulus, and sensitization is an increased response due to a repeated stimulus. An example of habituation is factory workers that already accustomed to the sound of machinery in the surroundings, and example of sensitization would be an annoyed person due to his neighbor keeps knocking on his wall for no reason. Habituation tends to be a stimulus-specific, and sensitization is rather a stimulus-general. Since the study of sensitization is rather lack of attention from researchers, we shall take a closer look into the study of habituation.

According to Thompson & Spencer (1966), habituation is defined as a behavioral response decrement that results from repeated stimulation and that does not involve sensory adaptation/sensory fatigue or motor fatigue. They also described nine characteristics of habituation. As of August 2007, a group of 15 researchers (Rankin et al., 2009) specialized in habituation in a wide variety of species, decided to refine the characteristics and added one additional characteristics. The revised characteristics becomes the following.

Characteristic #1 Repeated application of a stimulus results in a progressive decrease in some parameter of a response to an asymptotic level. This change may include decreases in frequency and/or magnitude of the response. In many cases, the decrement is exponential, but it may also be linear; in addition, a response may show facilitation prior to decrementing because of (or presumably derived from) a simultaneous process of sensitization.

Characteristic #2 If the stimulus is withheld after response decrement, the response recovers at least partially over the observation time (“spontaneous recovery”).

Characteristic #3 After multiple series of stimulus repetitions and spontaneous recoveries, the response decrement becomes successively more rapid and/or more pronounced (this phenomenon can be called potentiation of habituation).

Characteristic #4 Other things being equal, more frequent stimulation results in more rapid and/or more pronounced response decrement, and more rapid spontaneous recovery (if the decrement has reached asymptotic levels)

Characteristic #5 Within a stimulus modality, the less intense the stimulus, the more rapid and/or more pronounced the behavioral response decrement. Very intense stimuli may yield no significant observable response decrement.

Characteristic #6 The effects of repeated stimulation may continue to accumulate even after the response has reached an asymptotic level (which may or may not be zero, or no response). This effect of stimulation beyond asymptotic levels can alter subsequent behavior, for example, by delaying the onset of spontaneous recovery.

Characteristic #7 Within the same stimulus modality, the response decrement shows some stimulus specificity. To test for stimulus specificity/stimulus generalization, a second, novel stimulus is presented and a comparison is made between the changes in the responses to the habituated stimulus and the novel stimulus. In many paradigms (e.g. developmental studies of language acquisition) this test has been improperly termed a dishabituation test rather than a stimulus generalization test, its proper name.

Characteristic #8 Presentation of a different stimulus results in an increase of the decremented response to the original stimulus. This phenomenon is termed "dishabituation." It is important to note that the proper test for dishabituation is an increase in response to the original stimulus and not an increase in response to the dishabituating stimulus (see point #7 above). Indeed, the dishabituating stimulus by itself need not even trigger the response on its own.

Characteristic #9 Upon repeated application of the dishabituating stimulus, the amount of dishabituation produced decreases (this phenomenon can be called habituation of dishabituation).

Characteristic #10 Some stimulus repetition protocols may result in properties of the response decrement (e.g. more rapid rehabituation than baseline, smaller initial responses than baseline, smaller mean responses than baseline, less frequent responses than baseline) that last hours, days or weeks. This persistence of aspects of habituation is termed long-term habituation.

Readers are encouraged to check out the details, including evidence and reasoning behind the revised characteristics in Rankin et al. (2009). Although the ten characteristics clarifies further our concept of habituation, it seems that habituation, which is known to be "the simplest form of learning" is not so simple anymore. What Rankin et al. (2009) considered simple is the acquisition of habituation. There are complex mechanisms underlying the simple concept of habituation, deals with the nervous systems that constantly evaluating incoming stimuli and distinguish the stimuli which are important or not. This remains a challenge to the related researchers to figure out the details regarding the mechanisms.

3.5 Recent Models of Personal Memory Organization

The consensus from a wide range of studies views that there are two principle of human memory organization: temporal and thematic. Temporal organization deals with either chronological order of the experienced events or based on lifetime periods of one individual, and thematic organization deals with event-specific themes, such as holidays, family, or meetings. Robinson (1976) conducted an experiment about experience recollection involving activity (e.g., throwing) or an object (e.g., car) associated with an emotion word (e.g., happy). He evaluated the experiment through the memory recollection time, and found out that recalling with emotion as one of the cues slows down the recollection time. He then argues that people do not organized their memory based on emotions, because many different life experiences share the same emotions. However, Schulkind & Woldorf (2005) tried a different approach using musical cues to represent the underlying emotion, instead of using emotional word as the cue. They suspected that emotion words will was not best representing the emotional memories, as musical cues induce a particular mood to the subject as an emotion. After the recollection, they asked the participants to date the memories, rate the valence of emotions in the degree of positive to negative valence, and also arousal. What they found out was valence rating is higher elicited by positive music compared to the negative counterparts, and they concluded that emotion *does* have a role in the personal memory organization, and a better memory model needs to be considered.

3.6 Chapter Summary

Figure 3.8 depicts a more detail interconnection between Sensory Memory, Baddeley's updated WM model, and LTM storage. The chapter is intended to serve as a brief coverage of supporting evidence with respect to the core materials presented in the dissertation, instead of a comprehensive literature. Therefore, there are a lot more materials for the future prospect of this research to be covered in a deeper level that is not addressed here, such as forgetting, role of emotion in personal experience, memory reminiscence and abstraction, false memories, factors that influences memory recollection, just to name a few. In the next chapter, we discuss the implemented developmental robot architecture inspired from the insights and evidence discussed in this chapter.

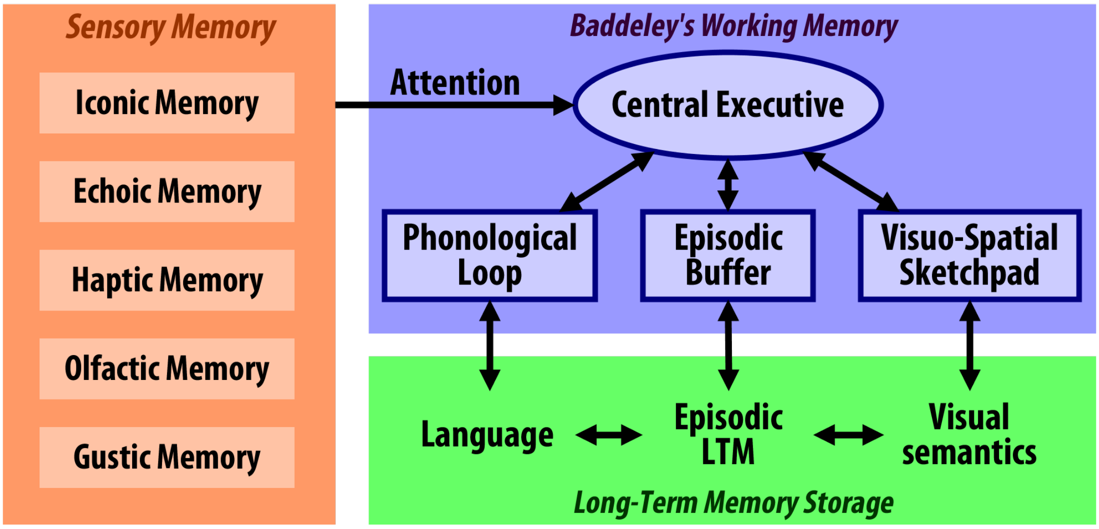


Figure 3.8: Component-wise relationship between Sensory Memory, Working Memory, and Long-Term Memory

Chapter 4

Interconnectivity of System Architecture

> No memory is ever alone; it's at the end of a trail of memories, a dozen trails that each have their own associations.
>
> Louis L'Amour

We introduce ERIS (Epigenetic Robot Intelligent System), inspired by current knowledge in human memory organization, as discussed in the previous chapter. In this chapter, the formalism of ERIS is elaborated and the interconnectivity between components, which is the main emphasis of the chapter, is evident. We begin the chapter by providing overview to the chapter, as well as literatures of relevant robotics application with respect to memory-based architecture that has been proposed in the past few years.

4.1 Overview

In order to acquire knowledge (and therefore *experience*), humans interact with each other and their environment to assess, confirm or revise their beliefs. As we posited in previous Sections, we argue that a similar process may be beneficial also for robots. In particular, we envisage a human-robot interaction scenario where robot knowledge goes through a human-assisted revision process, which is based on a labeling procedure.

In principle, robot knowledge can be classified as *raw* and *revised,* the latter being validated by a human during the interaction process. Whilst raw knowledge represents robot memory item initially consolidated during the interaction, revised knowledge originates as a result of the human-assisted assessment of raw knowledge. At any time, a human can provide robot knowledge with appropriate semantic labels, which constitute an assessment of initially acquired knowledge. As humans are able to confirm and revise their knowledge and beliefs multiple times, revised robot knowledge can still be subject to further revision. Such semantic labels (henceforth referred to also as *tags*) are then used by a robot when inquired about its own experience. When an inquiry is made by the interacting human, the information in the inquiry defines a context, which is used by the robot to frame its response. Terms in the inquiry are matched against tags associated with revised knowledge, a process we refer to as *familiarity mechanism based information retrieval*. On the basis of this mechanism, the robot can provide appropriate responses.

Although a precise characterization of the human memory architecture is still subject of active research, it is believed that three main components are present, namely Sensory Memory, Working Memory (previously known as Short-Term Memory) and Long-Term Memory. Sensory memory is responsible to handle the reception of the perceived stimulus by our five senses. While working memory is responsible for processing *active* information, long-term memory is considered as an almost *infinite* storage used to store memory item consolidated from working memory for a later use. Long-term memory consists of three sub-components, namely Semantic Memory (SM), Episodic Memory (EM), and Procedural Memory (PM). Each sub-component is responsible for different kinds of information: SM for storing facts and general knowledge about the environment, EM for storing experienced events related information, and PM for storing procedural information, such as motion commands or other behavioral skills. Information within SM and EM can be explicitly recalled as part of conceptualization processes whereas – usually – information within PM cannot.

In our human-robot interaction scenario, the robot is capable of gaining knowledge progressively from visual stimuli. Initially, memory items are consolidated as raw knowledge. This means the memory items are stored in Long-Term Memory and can be used by the robot, but they are not suitable to be used for human-robot interaction, since they must first go through the human-assisted knowledge revision process (which, in our case, mimics the progression towards mutual understanding). A human-computer interface has been implemented to facilitate knowledge revision, as the part of robot interaction module of ERIS.

4.1.1 Memory Models and Terminology

Albeit there is no widespread consensus about a general framework, memory models typically assume a multi-storage organization. Two models constitute fundamental milestones in the literature, namely the *multi store model* by Atkinson & Shiffrin (1968) and the *working memory model* by Baddeley & Hitch (1974).

Adopting a computational approach, the multi store model describes how information are formed into memory items and organized into different memory models. Three stores are usually identified, namely the Sensory Memory, the Short-Term Memory (STM) and the Long-Term Memory (LTM). Different processes are involved in the management of such an information flow. After being perceived and properly conveyed to the brain through relevant neural pathways, sensory information is represented inside Sensory Memory (in general, available for less than a minute). If sensory information is *attended*, the relevant part of it is transferred to STM, where it is processed for immediate use (occurring less than a minute). Then, if such a representation is *rehearsed* (an elaborative process further developed by Raaijmakers & Shiffrin (2003)), it is transferred to LTM (in principle, therein available forever). Otherwise, it is lost from STM according to a memory-trace decay process.

Baddeley & Hitch (1974) proposed a model for Short-Term Memory (which they call Working Memory - WM) that aims at better characterizing its subcomponents, each one devoted to represent and process different types of information. Specifically, WM consists of the Central Executive that orchestrates the behaviors of two subcomponents, namely the Visuospatial Sketchpad and the Phonological Loop. Central Executive is believed to deal with cognitive tasks related to logic and to make an on-demand use of subcomponents. Visuospatial Sketchpad processes visual and spatial based information, e.g., related to any motion in the environment. Phonological Loop deals with symbol-mediated information (i.e., which can be written or spoken), and can be further divided in two parts, namely the Phonological Store (linked to speech perception) and the Articulatory Control Process (linked to speech production) (see Jones, Macken, & Nicholls, 2004; Shaw & Tiggemann, 2004).

As a consequence of follow-up experiments, the original model has been updated by Baddeley (2000) to include a third subcomponent managed by the Central Executive, namely the Episodic Buffer. The role of the Episodic Buffer is to mediate between LTM and other components of WM: when WM is capable of identifying an observable relevant event (as a result of Visuospatial Sketchpad and Phonological Loop processing), the Episodic Buffer appropriately manages its storage in LTM. Nowadays, there is no shortage of reasons to believe that STM is made-up of a number of subcomponents. The WM model accounts for a number of real-world functional behaviors, such as task and verbal-level reasoning, reading and comprehension, problem solving, as well as visual and spatial information processing.

With respect to LTM, as proposed by Atkinson & Shiffrin (1968), two parts can be identified, i.e., explicit and implicit memory (Wood, Baxter, & Belpaeme, 2012). Explicit memory (also referred to as Declarative Memory) refers to consciously available memory items. It can be further divided in three subcomponents, namely the Episodic Memory (EM), the Semantic Memory (SM), and the Autobiographical Memory. EM is related to the encoding of generic events localized in time. An example of EM is the set of specific event occurred during the interaction with someone or with the environment. Knowledge about facts and their meaning is stored in SM. Differently from the content of EM, SM is not believed to depend on contextual information (Spaniol, Madden, & Voss, 2006). Autobiographical Memory is the knowledge related to both personal events and self-related information. However, it is noteworthy that Autobiographical Memory is different from EM, in that it refers to events strictly performed by the person, as pointed out by Conway & Pleydell-Pearce (2000) and Nelson & Fivush (2004). Finally, implicit memory (in particular, Procedural Memory) widely refers to motor action, specifically actions involved in the use of objects (including grasping, manipulation and tool use), as well as body motions (Bullemer, Nissen, & Willingham, 1989).

4.1.2 Models of Memory Components

Aside from the whole architecture literatures discussed in Chapter 2, a number of approaches are devoted to model specific memory components in robotics domain. With respect to Semantic Memory, two approaches are particularly interesting in our case, namely those put forward by Dodd (2005) and Dayoub et al. (2010).

The objective of the SM component designed by Dodd (2005) is to maintain information about objects located in the environment. This is achieved using a novel architecture combining the so-called Sensory EgoSphere later refined by Peters II, Hambuchen, & Bodenheimer (2009), as well as SM, WM and Central Executive. Although interesting, the framework is characterized by a number of drawbacks, as follows: (i) *a priori* knowledge about objects and the associated symbol grounding (Harnad, 1990) is required; (ii) since SM is designed to model and recognize objects in a very specific application domain, SM lacks the ability to represent anything that is not related to objects. In spite of these flaws, the framework has nonetheless the advantage of exhibiting a partial interconnectivity between the memory items pertaining to EM and SM.

Dayoub et al. (2010) proposed a SM component based on the multi store model of human memory advocated by Atkinson & Shiffrin (1968), specifically in the context of semantic mapping tasks carried out by a mobile robot. The robot is able to track the displacement of several objects using omni-directional vision and provide humans with the most likely suggestion about the location of any tracked object within the map. The overall behavior is managed using finite state machines. The advantages of the framework include: (i) a strong interconnection between the representation of objects, their locations and the capability of updating the internal model of the environment (i.e., the map); (ii) SM is tightly connected with the object tracking module, and it provides humans with comprehensive information about the map as a result of a human-robot interaction process. Specifically, humans may pose questions such as *Where was object* x *the last time you have seen it?* or *What are the most likely locations to find object* x *in the map?* Since robot knowledge is only limited to object properties and locations, the scope of questions that can be posed by humans is quite limited. However, the possibility of posing questions inspired us to implement a question-based knowledge information retrieval process.

As far as PM is concerned, the approaches by Salgado et al. (2012) and Dodd (2005) have been considered. The PM component by Salgado et al. (2012) stores basic skills and behaviors as a library to ground robot learning. Specifically, a Sony AIBO robot is expected to learn a ball catching behavior, strikingly similar to the work by Bellas et al. (2010). Whilst the architecture has been designed to implement adaptive learning techniques, it features a model of PM that turns out not to be consistent with state of the art of psychological studies. Furthermore, the information that can be obtained as a result of human-robot interaction processes is limited due to the inability of the system to store any information other than the learned associated behavior.

In the PM design proposed by Dodd (2005), robot motions are represented as nodes in a graph-like structure labeled as behavior nodes, motion primitive nodes, and example nodes. The architecture is designed to select PM nodes and to properly sequence them (Ratanaswasd, Gordon, & Dodd, 2005; Mastrogiovanni & Sgorbissa, 2013). Even though motions generated by sequencing robot behaviors are claimed to be fairly smooth, the PM design is highly dependent on the employed modular controller, as it as been pointed out by Ratanaswasd et al. (2005), the used behavior interpolator, and the trajectory error-reduction algorithm. Furthermore, since the structure of PM nodes only contains information about the associated behavior and the corresponding 3D trajectory, no comprehensive memory component interconnection is actually possible.

Finally, three approaches to model EM have been considered in our analysis, namely the work by Jockel et al. (2007); Stachowicz & Kruijff (2012) and, again, by Dodd (2005). Consistently with the notion of EM, the approach proposed by Stachowicz & Kruijff (2012) is focused on a formal framework used to represent and relate events occurring in both space and time into *spatiotemporal contexts*. In particular, a hierarchy of events is envisaged, where an *event* can be either *atomic* or *complex*. Atomic events can be combined in different ways to form so-called *sub*events and *super*events. Unfortunately, no formal account is provided about the adopted notion of context and – above all – its influence on the other components of the architecture. The proposed EM design also lacks any correlations between EM and Episodic Buffer, specifically in view of a continuous knowledge acquisition process while interacting with the environment.

The design for EM proposed by Jockel et al. (2007) assumes that an event is hierarchically classified as belonging to one of the following classes: perceptional event, command event, and executive event. In this case, an event is associated with procedural callback mechanisms. There are two among the claimed advantages of the architecture: the possibility of storing past experiences in a life-long memory storage component, and the ability to perform *one shot* learning processes. Again, no formal definition of such a notion of event is provided. This is surprising, given the argument that EM essentially consists of sequences of events.

Finally, the EM component designed by Dodd (2005) assumes it to be a medium for robot learning processes. Temporally sequenced records of specific events are stored as memory items called *episodes*. An association is maintained between EM items and the content of SM and WM, as well as task-related information (in a sense, mimicking the availability of PM). Episodes are retrieved from EM using an approach similar to what has been discussed by Anderson (1990) in the context of the ACT-R architecture. The main disadvantage of the approach is the difficulty of determining the *correctness* of a retrieved episode. The authors argue that this is due to the lack of a formal *context* definition. Nonetheless, our definition of EM is inspired by these design choices.

From the analysis of the literature, it emerges that two topics are fundamental to design a memory-inspired robot framework, namely a clear design of the architecture (including all its relevant components and interconnections) and an assessment about how contextual information impacts on memory items storage and retrieval.

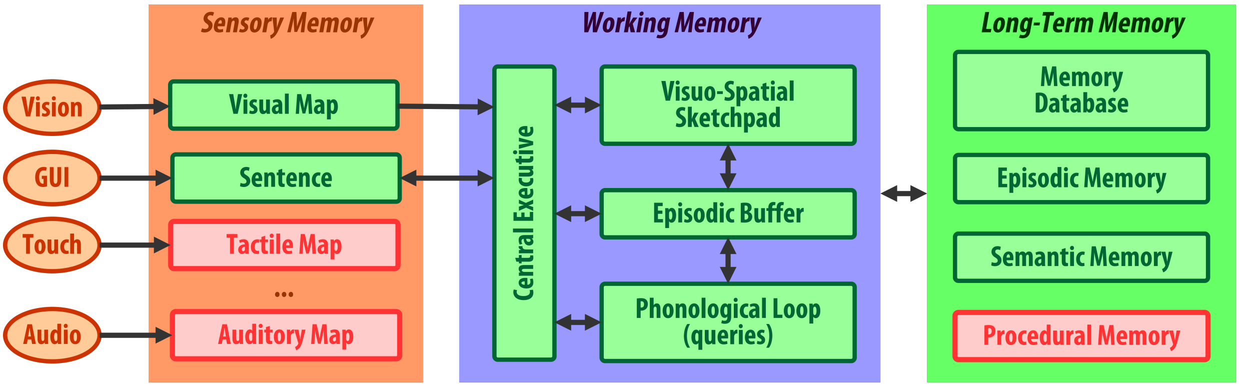


Figure 4.1: A graphical representation of the proposed memory architecture: parts in green corresponds to currently implemented components.

4.2 Connections with Memory Architecture

We introduce our proposed memory-inspired architecture called **Epigenetic Robot Intelligent System (ERIS)**, which the structure is outlined in Figure 4.1. ERIS is implemented as a ROS stack, and the general architectural design is inspired by the multi store model by Atkinson & Shiffrin (1968) updated with the WM model by Baddeley (2000). Each store can be further divided in subcomponents, according to the current understanding of memory organization in humans and other beings (Baddeley & Hitch, 1974; Wood et al., 2012). We assume the presence of a number of sensory components feeding different parts of the Sensory Memory. Currently, ERIS supports visual maps (in the form of *bitmaps*, but other approaches may be used as complimentary as well, for instance the framework by Antonelli et al. (2014)), and a simple mechanism to represent *questions* that can be posed to the system (as context-based cues), in a spirit similar to the work by Dayoub et al. (2010), as well as robot *answers* (as familiarity-based cues). Visual maps correspond to the basic representation used by the adopted vision algorithms. In principle, Sensory Memory can accommodate other sensory maps, such as tactile and auditory maps (Kallaluri, Even, Morales, Ishi, & Hagita, 2013; Denei et al., 2015).

In our current implementation, the visual map is manually segmented as a perceived scene from a continuous stream of visual feed, therefore the robot's and human's hand are not captured within the scene. The visual map is always transferred to be processed in WM within the Visuospatial Sketchpad. Relevant changes in the perceived visual feed constitute *scenes*. The identification of scenes is related to the formation of EM memory items (called *episodes*) inside the Episodic Buffer. This process is managed by a proper computational component representing the Visuospatial Sketchpad, which we call Visual

Stimuli Processor (ViSor). Inside ViSor, visual maps are processed using two feature extraction techniques, color and texture features, for recognizing the detected objects. The color feature is using Color Structure Descriptor, and the texture feature is using Edge Histogram Descriptor. Both descriptors are part of the MPEG-7 standard (Manjunath, Salembier, & Sikora, 2002). In other words, a specific image representing the changes within the environment is defined as a *scene*, and scenes are used to form an *episode*. Multiple episodes are encoded as collections of EM items. An *event* consists of several episodes sequentially ordered based on timestamps.

Each memory item is modeled as a collection of cue-value pairs, where cues correspond to features extracted from incoming images. Once the scene has been captured and processed through the ViSor, an episode is formed and consolidated into the LTM storage. Here, a scene is captured after each pick and place movement has been performed by the robot. Once a scene has been captured and processed through the ViSor component, an episode is formed and consolidated into the LTM storage. Saliency detection has been considered in the proposed framework given the widespread belief that it plays a central role in the human memory consolidation process and episodic segmentation (Posner & Petersen, 1990; Kaster & Ungerleider, 2000; Jeong, Arie, Lee, & Tani, 2011; Posner & Petersen, 2012). Schillaci, Bodiroža, & Hafner (2013) provides an excellent analysis of the influence of saliency in a human-robot interaction domain.

Currently, only EM and SM have been implemented within LTM, and cue-value pairs in LTM are represented using a relational database. Relevant results (i.e., episodes or SM items) temporarily stored in the Episodic Buffer are compared with the Memory Database, which keeps track of familiar SM items and episodes, and consolidated when either they are not familiar or not listed in the database. The bidirectional arrow connecting the Episodic Buffer (specifically, the ViSor module) and LTM (specifically, the Memory Database) in Figure 4.1 represents the ability to consolidate and recall memory items. As postulated by Eichenbaum & Cohen (2001) and Tulving (2001, 2002), human memory is characterized by the property of undergoing a continuous, subjective rehearsal and active modification, which is how we actually re-experience past events during memory recollection. Even though a precise understanding of this phenomenon is still subject to research efforts, our framework aims at mimicking this feature of the human memory, which without any doubt plays a central role in everyday behavior.

From a computational point of view, the choice of which information to store inside LTM as a collection of cue-value pairs is an important design parameter for the whole architecture. It is necessary to find a trade-off between the proper selection of image features (i.e., to be stored as cues) best discriminating among different episodes (i.e., having well-separable value spaces), and the need for storing the minimum amount of information (i.e., the size of the LTM storage) given the continuous nature of the knowledge acquisition process. Two main ideas are considered for this matter: (i) considering that every computer science problem is related to the famous "*time-space trade-off*" regardless of the capacities and availability of computer memories in the present and the future; and (ii) anticipating the increasing needs of storing more information in a single memory item in the future (compared with our currently implemented color and shape information). In particular, although the capacity of computer memory is considered abundant and inexpensive nowadays, having succinct representation of memory items allows for more efficient memory retrieval processes, which eventually allows more memory items to be stored, as well as boosts runtime performance.

A similar information flow can be determined when the user asks the robot to recall previously acquired memory items. Currently, this is done using the cue-value pair based formalism that is mapped to specific queries in the Phonological Loop to be submitted to LTM. The same cue-value based formalism is used to present to a human user the robot accounts related to what has been actually recalled. It is noteworthy that these two information flows are not to be considered in strict alternative. In fact, it is possible to pose questions while the robot is still acquiring new knowledge.

The ability to manage different parts of STM is due to our implementation of Central Executive. In human memory, the Central Executive is believed to be responsible for processing information originating from different sources, coordinating a number of otherwise passive subsystems, as well as performing selective attention and inhibition strategies (Baddeley, 1996, 1998; Collette & Van der Linden, 2002). In the current implementation, Central Executive is designed as a computational process able to perform a number of tasks, as follows:

1. Managing the encoding processes of Episodic Buffer to store relevant visual information computed by Visuospatial Sketchpad (e.g., object shapes, colors or locations as perceived in a scene) in the form of cue-value pairs in such LTM components as EM and SM.

2. Performing familiarity-based information retrieval, i.e., identify relevant cues, based on logical processes involving cue analysis and problem awareness (Mastrogiovanni, Scalmato, Sgorbissa, & Zaccaria, 2011; Mastrogiovanni & Sgorbissa, 2012).

3. Executing recollection processes, i.e., recalling memory items from LTM using the

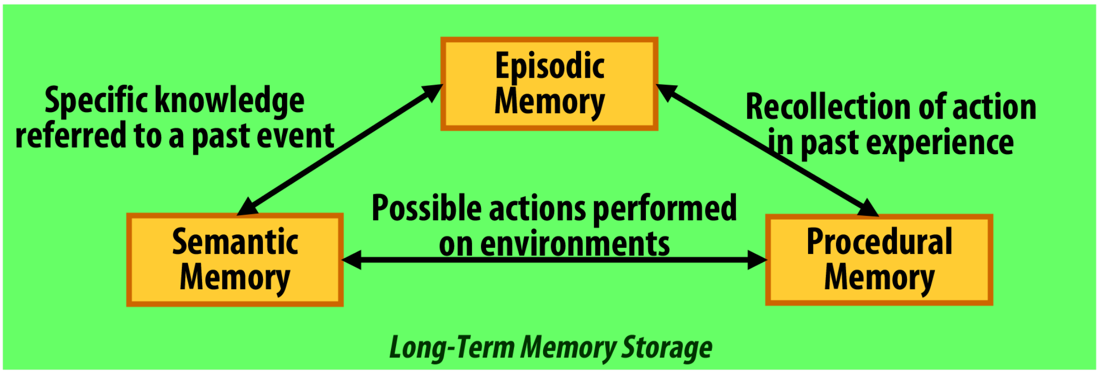

Episodic Memory
Specific knowledge referred to a past event
Recollection of action in past experience
Semantic Memory
Possible actions performed on environments
Procedural Memory
Long-Term Memory Storage

Figure 4.2: Interconnectivity between Semantic-Episodic-Procedural Memory in terms of past experience.

results of the familiarity-based retrieval process.

4. Supervising the Phonological Loop to analyze cue-value pairs based information related to recalled LTM memory items.

4.3 Formalism

In this Section, we define the most important concepts of the proposed architecture (ERIS), thereby defining the memory model upon which the framework is designed and implemented. We introduce first the notion of *memory item*. We will later use the definition of memory item to formally define elements in Semantic Memory and Episodic Memory.

Definition 1 (Memory Item) *A Memory Item $i \in I$ is a set of n cue-value pairs, such that $i = \{(c_1, v_1), \ldots, (c_n, v_n)\}$.*

A memory item is a single element that can be used to represent any of the subcomponents of Long-term Memory, such as Semantic Memory, Episodic Memory or Procedural Memory. Here, we do not model Procedural Memory. However, it is noteworthy that we explicitly take into account the link between the knowledge represented in Semantic Memory and Episodic Memory (see Squire & Kandel, 2000). As we discussed in Chapter 3 and Section 4.1.1, Semantic Memory stores general-purpose knowledge about the environment in terms of concepts and their relationships (which are, in a sense, independent from the particular robot and therefore transferable to other robots), whereas Episodic Memory represents robot experiences (in the form of episodes) anchored to a specific point in space and time (which is typically robot-dependent). Figure 4.2 shows the interconnectivity of Long-Term Memory components in terms of past experience.

Definition 2 (Entity) *An Entity ϵ is a grounded memory item $i_\epsilon \in E$, with $E \subset I$.*

Entities are a representation of objects in the environment, humans and other agents acting therein. Each entity is mapped to a set of grounded cue-value pairs, where the semantics associated to cues globally define the entity as a type.

Definition 3 (Object) *An Object n is a grounded memory item $i_n \in N$, with $N \subset I$, where i_n is defined in terms of three multi-valued cues, i.e., name, shape and color, and by a number of Boolean cues, i.e., graspable and manipulable.*

Each memory item corresponding to an object is characterized by specific values associated with its constituent cues.

Definition 4 (Location) *A Location l is a grounded memory item $i_l \in L$, with $L \subset I$, where i_l is defined in terms of one numerical cue corresponding to a 2-element vector pos2d and one Boolean cue type.*

A memory item representing a location can refer to either an *absolute* or *relative* 2D position (expressed using the type and pos2d cues, respectively), whose semantics depends on the specific Cartesian frame with respect to which the location is expressed. For example, the description of the previously introduced *bluebox* object can be augmented with a description $\{(\mathsf{pos2d}, (72, 13)), (\mathsf{type}, relative)\}$.

We also introduce a notion of time inspired by a simple linear time logics approach (Emerson & Halpern, 1986), as follows.

Definition 5 (Time Instant) *A Time Instant t is a cue-value pair, with $t = (\mathsf{time}, integer)$.*

Time instants are represented in Unix epoch time, which are positive integer numbers. In ROS structure of memory item, time instant belongs to header cue in each message.

Definition 6 (Semantic Memory) *A Semantic Memory SM is a collection of k grounded memory items $\{i_1, \ldots, i_k\}$, which can be divided into 5 disjoint sets, such that $SM = \{N, H, L, T, W\}$, where: N represents known (or previously identified) objects, H stores information about humans or other agents the robot interacts with, L is related to entities spatial information (locations), T represents entities temporal information (time instants), whereas W is an association between lexical knowledge and entities.*

We separately model N and H in order to account for inanimate objects and intentional agents, respectively. As previously noted, in this dissertation we focus on the set N and not on H, which is characterized by the appropriate knowledge to model the objects the robot interacts with.

The representation of object is first consolidated as an Semantic Memory item whenever a novel object is detected by the ViSor module, through a process which resembles *habituation*, which literature has been explained in Chapter 3.4.7.

Definition 7 (Episode and Scene) *An Episode $\hat{\sigma}$ is a memory item succinctly representing the captured visual changes of the environment, which is a collection of b grounded memory items $\{i_{\hat{\sigma},1}, \ldots, i_{\hat{\sigma},b}\}$, which occur at a time instant t_σ. The visual change occurring at a time instant t_σ is defined as a scene, which is a sequence of visual feed $\{\sigma_1, \ldots, \sigma_b\}$.*

In particular, a scene is an instance of a captured image in the visual stream, which is then later processed using the ViSor module and yields saliency information. The saliency information is further used to determine the familiarity status of the detected scene. Novel scenes are visually processed by object recognition module, which results contributes to a formation of **an episode (a scene representation, and also an EM item)**. The object recognition module includes the color and texture feature extraction method. Currently, familiar scenes will not be processed further. Episodes can employ both memory items related to objects represented therein as well as global descriptors of a scene, such as the number of objects, through the cue count. Two subsequent scenes are separated by a significant change in the image saliency level.

Definition 8 (Event Type) *An Event Type $\xi = (\textsf{type, active/passive})$ is a cue-value pair consists of either active or passive event.*

Events are classified as being active or passive. An active event originates from one or more actions performed by the robot itself, whereas a passive event either corresponds to actions carried out by humans interacting with the robot or to something that simply happens in the robot workspace and is perceived in a scene. In this dissertation, we consider both active and passive events. Although the robot witnesses events that are influenced by its own motions, it should be noted that since the Procedural Memory is not considered at the moment, active events will not influence conducted experiments.

Definition 9 (Event) *An Event η is a collection of s episodes $\{\hat{\sigma}_{\eta,1}, \ldots, \hat{\sigma}_{\eta,s}\}$, with associated type information.*

An event is defined by two corresponding initial and final scenes (represented as episodes), namely $\hat{\sigma}_{\eta,1}$ and $\hat{\sigma}_{\eta,s}$, as well as by all intermediate scenes. Our concept of event has three interesting properties: flexible, subjective, and personal. These three properties will become apparent in the experiment later, and the details are discussed in the beginning of Chapter 6.

Definition 10 (Episodic Memory) *An Episodic Memory EM is a collection of z events $\{\eta_1, \ldots, \eta_z\}$.*

The knowledge retrieval process is based on the notion of context and context item.

Definition 11 (Context and context item) *A context ξ is modeled as a set of X context items γ_ξ, such that $\xi = \{\gamma_{\xi,1}, \ldots, \gamma_{\xi,X}\}$. A context item γ_j is characterized by a **retrieval cue** c_j, a **value** for that cue v_j, and a set of relevant **tags** Ψ_j, such that $\gamma_j = \{c_j, v_j, \Psi_j | 1 < j < X\}$.*

Differently from memory items, contexts are not part of the set I, meaning that they do not necessarily correspond to definitions of entities, objects or locations. In our framework, contexts are used in the knowledge retrieval process to recall memory items stored in the Long-term Memory. As it will be discussed in Section 6.3, humans interacting with the robot can pose a number of questions, which are formally encoded as contexts.

To this aim, cues can be classified as general-purpose and context-dependent, depending on their memory scope. For instance, cues may be appropriate to all the available memory components (e.g., Semantic Memory and Episodic Memory), or be related to one component exclusively (e.g., Semantic Memory only).

Definition 12 (General-purpose Cue) *A cue c is general-purpose if it refers to a memory item i that is not specific to any memory component.*

A context using a general-purpose cue may include, for instance, information related to both Semantic Memory and Episodic Memory.

Definition 13 (Context-dependent Cue) *A cue c is context-dependent if it refers to a memory item i that is specific to a particular scene observed by the robot.*

As we discussed in Section 4.1.1, the visual stream in processed by Visuospatial Sketchpad (represented by the implemented ViSor module) to form episodes, which are consolidated in LTM as part of EM through the Episodic Buffer. Context-dependent cues are used to retrieve memory items stored in EM.

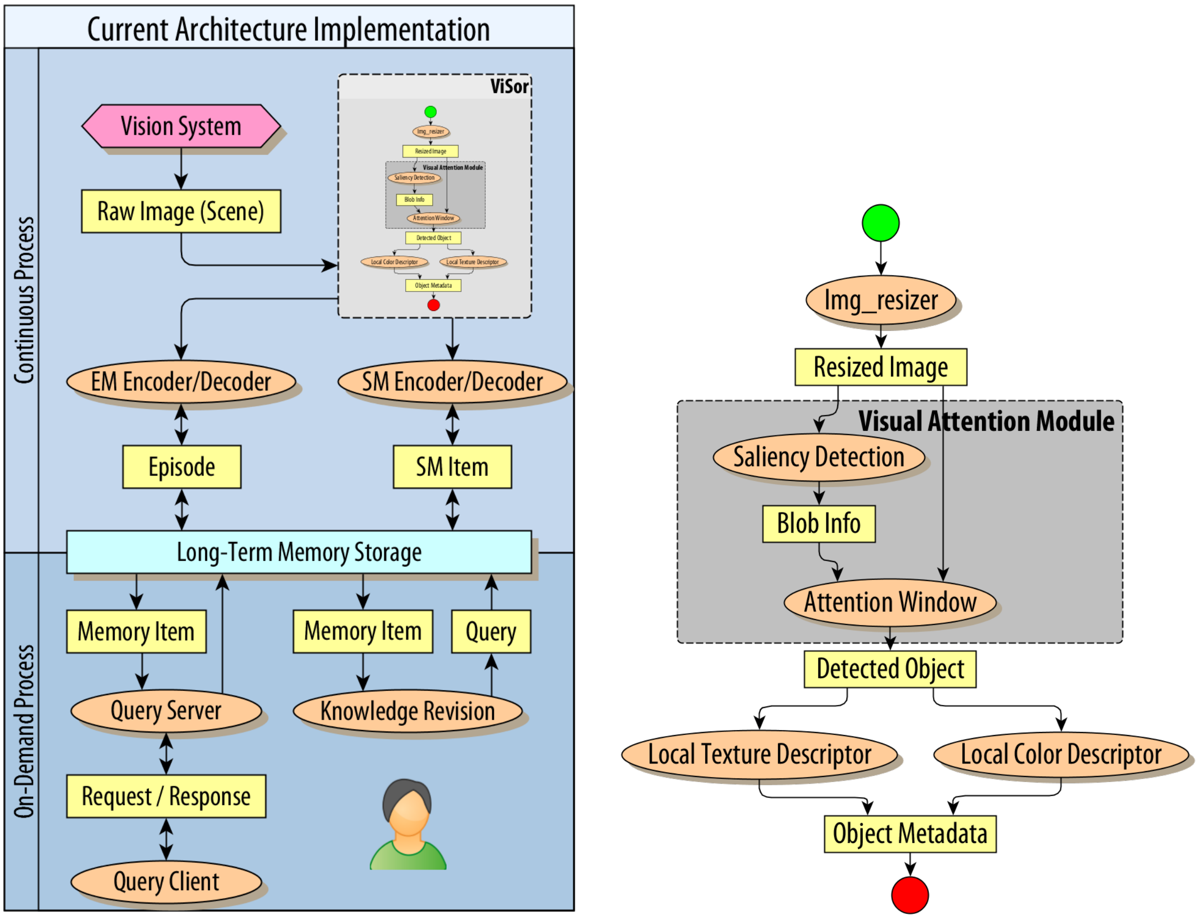
In a complete sensorimotor process, it is believed that the consolidation process involves Semantic Memory, Episodic Memory and Procedural Memory, as discussed by Tulving (1985) and Squire (2004).

4.4 Implementation

The implementation of ERIS[1]takes a full advantage of the features offered by ROS (Quigley et al., 2009), allowing easy expansion and accessibility to general-purpose robots. There are two ROS packages implemented so far, (1) the Visual Stimuli Processor (ViSor); and (2) the core architecture (ERIS). Figure 4.3a depicts the two main process available within the current ERIS core architecture implementation including the ViSor package. The ovals represent active elements (i.e., ROS nodes), the squares represent passive elements (i.e., message passed in between nodes and LTM storage), and the arrows represent the information flow between nodes. The upper half of the diagram represents a continuous process of visual-based developmental knowledge acquisition. The ViSor package deals with raw image and pass the results to ERIS module to be processed further, i.e., recall the existing memory element and consolidate a new one. Meanwhile, the lower half of the diagram represents an independent process of the human-robot interaction through the query client node to provide robot with the query and display the responses, and with a server node to process the query, as well as the ROS node that allows the knowledge revision process.

The architecture is fully functional regardless of the absence of Procedural Memory, which responsible for storing memories related to robot motor skills as it is highly dependent of its physical capabilities. Yet, it has lots of future potentials to be discovered from the developmental/epigenetic robotics perspective and extensible by other researchers. As for now, the robot movements does not involve the Procedural Memory. The architecture allows the robot to autonomously gather and develop knowledge based on visual stimuli, and interact with human based on Figure 4.3a. Now we discuss the ViSor package as it is responsible for all the image processing before dealing with the memory processing in the ERIS package.

[1] The code is publicly available at `https://github.com/ferdianap/eris`



(a) ERIS core package

(b) ViSor package

Figure 4.3: Simplified ROS diagram of current architecture implementation

4.4.1 ViSor Package

The ViSor package consists of global and local processes, as seen in Figure 4.3b. The global process covers the whole scene (at the scene level), and the local process covers each of the objects in a single scene (at the object level). The global process currently includes saliency-based object detection, and the local process includes color and texture feature extractors. The object detection module is included in the visual attention node, and both the color and texture feature extractors are implemented based on MPEG-7 (Manjunath et al., 2002) Color Structure Descriptor and Edge Histogram Descriptor, respectively. The Edge Histogram Descriptor covers texture features within an object, as well as the shape outline of 2D projection of the object with respect to the Field of View plane. To be concise, this package consists the minimal required modules to process the scene, distinguish and separate the contents of a memory element.

First, the image resizer node is responsible for down-sampling the scene for efficiency and performance reason into a resized image, which will be fed into both saliency detection node and attention window node. Then the saliency detection node yields each information about the detected object within the given scene (i.e., the metadata of the scene, position of each detected object), which directly processed with the corresponding scene. Then we get each of the object image with their corresponding information as well as the scene metadata. This leads to the color and texture local processing nodes, which yields object metadata ROS message, containing the texture and color representation for each object and their original object image. The object metadata here refers to the color and texture representation (descriptors), and the information shared here are the color and texture representation for each object and their original object image. The reason we stored the object image given that we have the representation is for two reasons: complies with *The Dual-Coding Theory* (see Sternberg, 2003, for more details), and for the knowledge revision, so that we can still get feedbacks from human based on the original image.

First, the captured scene in the form of raw image is being down-sampled into 160×120 pixel resolution for computational performance efficiency in the img_resizer_node. Then, the channels of the resized image are split into BGR and HSV and published into /processed_img/bgr8_img and /processed_img/hsv_img topic, respectively, by the img_chnldvdr node. In principle, this node provided for the convenience of image processing methods based on different color channels. In this case, we use only the BGR channels for a basic image processing of the scenes. The BGR image is then processed in the saliency_det_node for object detection based on saliency. The saliency_det_node is implemented as a regular method callback (for storing the current scene) and service callback (for processing the current scene on-demand). After processing the scene, the node publishes the corresponding blobInfo message, which will then later be processed in the attention window node attn_window_node. The node subscribes to both /processed_img/bgr8_img and /blob_info topics, crops the object blobs based on the information from the blobInfo message, and publishes the detected entity in an array form called detEntityArray. We have implemented object_extractor class for the convenience of object extraction within the attn_window_node. The detEntityArray is then process in two nodes in parallel as part of the local scene processing: local_color_node and local_shape_node.

The corresponding feature extraction methods are performed in each node, i.e., Color Structure Descriptor (CSD) in the local_color_node and Edge Histogram Descriptor (EHD) in the local_shape_node. The local_color_node publishes /local_feat/color topic, and the local_shape node publishes /local_feat/shape. These topics are the object metadata depicted in Figure 4.3b. In ROS, we implemented textureInfo.msg as the texture information of each detected object, where 80 is the default size of the Edge Histogram Descriptor (EHD), and the desc variable holds the corresponding vector values.

Listing 4.1: textureInfo.msg

```
## This is the texture description for each object
## By default, EHD is used.
## Related pkg: visor
Header header
int16 size # DEFAULT= 80 (VECTOR SIZE)
int16[] desc # EHD is used
```

For the colorInfo.msg, the default vector size for the Color Structure Descriptor (CSD) is 64. It should be noted that for both for textureInfo.msg and colorInfo.msg, the feature extraction methods and the selected vector size must be remain the same for the memory recollection in terms of familiarity, otherwise error will occurs due to the difference of the vector size and the feature similarity comparison will not make any sense due to the difference of features extraction methods.

Listing 4.2: colorInfo.msg

```
## This is the color description for each object
## By default, CSD is used.
## Related pkg: visor
Header header
```

```
int16 size # DEFAULT= 64 (VECTOR SIZE)
int16[] desc # CSD is used
```

We implemented ROS messages for both color and texture detected for each scene as localTexture.msg and localColor.msg.

Listing 4.3: localColor.msg

```
## This is the color description for each scene
## Related pkg: visor
Header header
visor/colorInfo[] object_color
int16 object_count
```

The localColor.msg contains the each of the object color detected within a scene and object count as an int16 type. While for the localTexture.msg contains several information such as object count, a set of array of object texture corresponding for each detected object, relative spatial information with respect to the field of view in the form of pos_minx, pos_maxx, pos_miny, and pos_maxy, as well as the corresponding image of the objects.

Listing 4.4: localTexture.msg

```
## This is the texture description for each scene
## Related pkg: visor
Header header
int16 count # obj count
textureInfo[] object_texture
int16[] pos_minx
int16[] pos_maxx
int16[] pos_miny
int16[] pos_maxy
sensor_msgs/Image[] image
```

The blobInfo.msg contains the relative spatial information of the detected objects in the form of rectangular blobs for each scene, and count as the amount of detected objects.

Listing 4.5: blobInfo.msg

```
## This is the blob information
## Related pkg: visor
Header header
```

```
int16 count
int16[] minx
int16[] maxx
int16[] miny
int16[] maxy
```

These information within the blobInfo.msg are then combined with the corresponding scene image in the attn_window_node, and assigned as the detEntityArray.msg.

Listing 4.6: detEntityArray.msg

```
## This is the detected entities in an array structure
## Related pkg: visor
Header header
int16 count
sensor_msgs/Image[] entity
int16[] pos_minx
int16[] pos_maxx
int16[] pos_miny
int16[] pos_maxy
```

4.4.2 ERIS package

Continuing from the previous subsection, the object metadata ROS message that we obtain from the ViSor package is fed into both EM and SM encoder/decoder nodes, which resulting in an episode and an SM item, respectively. The bidirectional arrows implies a checking condition, that if one episode or SM item is already exist in the Long-Term Memory storage, it will not be reconsolidated. The episode and SM item is independently processed, meaning that if an episode is already exist (a familiar scene is detected), the episode and all the SM item will not be consolidated. However, if one SM item is already exist (a familiar object is detected) but a novel scene is detected, the episode and other detected novel objects will be consolidated and only the SM item correspond to that familiar object will not be consolidated.

Now we move on to the human-robot interaction based on the lower half of Figure 4.3a. The left hand side represents the general interaction with human, where human can ask Baxter questions, in the form of lexical inputs from the query client node (currently implemented as a command-line interface), regarding its past experience. The question asked can be optionally paired with visual stimuli by letting the Baxter witness the current scene (e.g., the workspace arrangement or a single object) when posing the question. The right hand side correspond to the knowledge revision process, which has been detailed in Figure 5.5. The knowledge representation is explicitly represented by ROS message.

The package mainly consists of encoder module and HRI module. As the encoder module, an encoder_node has been implemented which currently applies for both SM and EM. The node subscribes to the /local_feat/shape and /local_feat/color, and although the node is responsible for memory formation and consolidation, it also performs memory recollection and familiarity checks of objects and scenes (Case 1 as elaborated in the later chapters of the fundamental cases of memory recollection).

During the memory recollection, we consider two kinds of familiarity: object and scene familiarity. An object familiarity is currently computed using two criteria: when both color and shape similarity value is over the threshold value. During the input familiarity of case 1, both color and texture familiarity check is done by cosine similarity and the threshold for both features is set to be 70%, hence above the threshold value would be considered as familiar. Cosine similarity is listed in Equation 4.1, where A and B are two different 1D feature vectors, applies to both color and shape features, assuming that A and B are the same type of feature and have the same vector size.

$$similarity = cos(\theta) = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{4.1}$$

Assuming that i_{nF} is the object that is considered familiar, the condition of object familiarity is listed in Equation 4.2, given two sets of 1D vector A and B for both color and shape features, as well as th_{color} and th_{shape} corresponds to the threshold value of color and shape features.

$$\begin{aligned} i_{nF} = &\{cos(\theta_{color}) > th_{color} \wedge cos(\theta_{shape}) > th_{shape} | \\ &A_{shape}, B_{shape}, A_{color}, B_{color}, th_{shape} = 70\%, th_{color} = 70\%\} \end{aligned} \tag{4.2}$$

This can still be improved in the future development of the project using, for instance, a robust fuzzy logic controller, considering similarity result such as 69% will be considered as “not familiar”. It is important to note that the HRI module here serves only as a means to analyze if the robot has developed and able to provide the human some information related to the recalled past experience.

On the other hand, the scene familiarity is determined by the object familiarity and the relative position of all detected objects. A scene is considered to be familiar when all the detected objects are considered to be familiar, and the relative position with respect to the field of view does not exceed the threshold pixel value, which is set to be 20 pixels.

If we consider σ_F as a familiar scene, i_{nF} as a familiar object (as detailed in Equation 4.2), and the threshold value th_{lx} and th_{ly} of 20 pixels, corresponds to both x-axis and y-axis of the field of view respectively, for checking the location of the detected objects, the formal notation of scene familiarity is detailed in Equation 4.3. Similarly to the object familiarity, a more flexible familiarity measures can be achieved by integrating a complementary fuzzy logic controller or a similar one.

$$\begin{aligned} \sigma_F =& \{\forall i_n \subset N, \forall l_x, l_y \subset L, (|l_{x\hat{\sigma}} - l_x| < th_{lx} \wedge |l_{y\hat{\sigma}} - l_y| < th_{ly})| \\ & i_n \in \hat{\sigma}, i_n = i_{nF}, l_x, l_y \in \sigma \wedge (l_{x\hat{\sigma}}, l_{y\hat{\sigma}}) \in \hat{\sigma}, th_{lx} = 20 pixels, th_{ly} = 20 pixels\} \end{aligned} \tag{4.3}$$

The structure of an episode is implemented as a ROS message, where it contains information such the time instant (stored in the header variable), episode label, obj_count as the number of detected object, seq_count as the number of performed motor skills, relative position of the objects detected in that particular episode, obj_name as the reference to the detected object name, and sequence as an array of performed motor skills (which is currently empty, since it is related to the PM).

Listing 4.7: episode.msg

```
## This is the structure of an episode within the Episodic Memory
## Related pkg: eris
Header header
string label
int16 obj_count
int16 seq_count
int16[] pos_minx
int16[] pos_maxx
int16[] pos_miny
int16[] pos_maxy
string[] obj_name #filename
string[] sequence
```

For the structure of a SM item, smEntity.msg contains information such as the encoding time instant (in the header variable), object label, an array of related tags, the image of the corresponding object, and boolean parameters (i.e., graspable and manipulable).

Listing 4.8: smEntity.msg

```
## This is the structure of a Semantic Memory,
## specifically about a single object only.
## Related pkg: eris
Header header
string label
string[] tag
sensor_msgs/Image image
int16[] color # Color structure Descriptor
int16[] texture # Edge Histogram Descriptor
#bool graspable
#bool manipulable
```

To track all the revised knowledge, we design the structure of a Familiarity Filtering Index (FFI) database ROS message file, such that the generated database file _FFI.db contains an array of known_objects and known_tags. Raw knowledge will not be listed in the database file, since no interaction context (in form of label and tags) are associated with the corresponding object. The database will be dynamically updated when there is a revised or removal of information, such as tags and object name, based on the interaction with human.

Listing 4.9: ffiDatabase.msg

```
## This is the structure of an FFI database
## Related pkg: eris
string[] known_objects
string[] known_tags
```

To manage all the known tags, once a tag has been associated to an object in an SM item, a tag file will be generated with the structure based on the tag.msg file, which contains the tag name and an array of the associated object.

Listing 4.10: tag.msg

```
## This is the structure of a Tag
## Related pkg: eris
string name
string[] assoc_obj
```

To allow the human to interact with ERIS using contextual information, we provided a command-line interface in a ROS server-client architecture. The query from human will be passed as the format according to query.srv, where the request consists of both interaction case and question in the form of int8 type as well as the array of string of the requested context. While the response is the multi_responses, which simply an array of string, for a flexible answer in a natural language by ERIS.

Listing 4.11: query.srv

```
## This is the parameters related to query provided by human
## Related pkg: eris
int8 icase # Interaction case
int8 question
string[] lexicalinfo
---
string[] multi_responses
```

Listing 4.12 shows the help instruction of the command-line interface, including on how to ask the verbal question in the different cases of memory recollection. For an experiment where the robot is not present, we have implemented the imageloader node to load a JPG image in a predefined directory to replace the role of capturing the live visual stimuli during the human-robot interaction. The image within the preset directory will be loaded as if the robot experience a live visual stimuli, and will be processed in the same fashion as in the real robot, in this case, valid for only case 1 and 3 of memory recollection.

Listing 4.12: Snippet of command-line help instruction

```
Universal Client Program ver. 0.1.0
Basic Command Line Parameter App
Example query (case 3, question 1, with context):
rosrun eris query_client -c 3 -q 1 -l leftmost -l lamp
Valid Case:
  1. Image only
  2. Context only
  3. Image+Context
Make sure that the `vision.JPG` image is located in
```

```
the folder 'BaxterVision'. (This serves as an alternative
for live visual stimuli during HRI.)
Choose question to ask Baxter:
<img> = valid for Case 1 & 3,
<ctx> = valid for Case 2,
<img/ctx> = valid for all cases.
1. Are you familiar w/ <img/ctx>?
2. What kind of <ctx> were you presented with?
   (result combined w/ q1)
3. How many of <ctx> have you seen so far?
4. Did you remove any of <ctx>?
5. Did you move any of <ctx>?
6. How many objects left after you remove <ctx>?
7. What <ctx> object did you move?
   (result combined w/ q5)
8. How many objects at least were in the workspace?
```

In principle, the process of memory recollection through human-robot interaction can be performed at any time, even in the middle of the progressive knowledge development process. This is because it is only dependent of the currently possessed knowledge instead of the knowledge development process itself.

In general, the architecture is very flexible for future expansion as well as addition of a more sophisticated robot development modules, and any ROS compatible robots are easily integrable.

4.5 Chapter Summary

This chapter provides the main contribution of the thesis, which is the architectural design of developmental framework in explicit formal notation. The research is motivated by the significant aspects that are currently missing in the state-of-the-art literature of the related research field (detailed in Section 4.1, also briefly discussed in Chapter 2) which are mainly about the development of cognitive architecture for various application domain, and including the fact that we need a clear distinction between cognitive and developmental robot architecture. The idea behind the research work is inspired based on human memory organization (discussed in Chapter 3), and the key building block of our developmental architectural design is the notion of **memory item** (discussed in Section 4.3). We have discussed about what has been achieved in terms of design overview (in Section 4.2) and implementation (in Section 4.4). In our current development stage, there are some elements that are not implemented yet, such as procedural memory, forgetting mechanism, and sensory elements other than visual. Also, we have not yet formalize the role of autobiographical memory, due to all the knowledge gathered by the robot are related to itself, instead of related to both general abstracted memory and self-memory. Therefore this can be considered also as the role of autobiographical memory, which function is significant to the developing system.

Chapter 5

Personal Knowledge Development

> We have associations to things. We have, you know, we have associations to tables and to - and to dogs and to cats and to Harvard professors, and that's the way the mind works. It's an association machine.
>
> Daniel Kahneman

This chapter provides the explanation of how a robot acquires and develops its own knowledge by interacting with human and the environment, based on the architectural design elaborated in the previous chapter. The chapter begins by providing an overview of how the familiarity mechanism works as well as the related literature and motivation, and the details of the memory processing related to familiarity mechanism, the notion of *context* and *tags*.

5.1 Overview

Association and classification is used by humans at a fundamental level (both consciously and unconsciously) to organize information. In this chapter, we describe how the human-assisted knowledge revision process exploits labeling to associate robot knowledge with tags, which are then used to implement a simple form of classification. As part of the human-robot interaction process, the robot interacts with the human to classify its factual and sensory knowledge. Tags are *dynamic*, since they are obtained as a result of human knowledge, and can be further revised in later interaction processes. The following two sections describe the literature and motivation of using the familiarity mechanism, and how such a semantic structure is used to implement a robot-based familiarity process, i.e., an associative process that allows a robot to retrieve relevant memory item in human-robot interaction processes.

The main motivations for a simple and generic classification procedure are: (i) familiarity measures based only on object features such as color and shape are highly subjective and prone to sensing errors, whereas more general semantic information, when agreed by both humans and robots, is stable; (ii) as discussed in Section 2.5, a knowledge revision process based on semantic tags allows for the definition of a shared context between the interacting human and robot.

The detected objects will undergo image processing module to have the color and texture features extracted as the familiarity determination condition. Color Structure Descriptor (CSD) and Edge Histogram Descriptor (EHD) are used for both color and texture/shape outline of the object, respectively. CSD corresponds to a representation of the color distribution and the local spatial structure of the color of an image. Formally, CSD is a 1D array of eight big-quantized values

$$CSD = \bar{h}_s(m), m \in \{1, \ldots, M\}$$

where M is the size of the features of either $\{256, 128, 64, 32\}$, and s is the scale of the associated square structuring element. In this case, we use the value of $M = 64$, as it is accurate enough to represent the feature, and the scale s is computed automatically depending on the image size. EHD computes the local edge distribution of the image by dividing the image space into 4×4 sub-images. By doing that, edges are categorized into 5 different types based on the angle, such as horizontal, vertical, 45° angle, 135° angle, and non directional edges.

Each feature is represented as a 1D-vector which can be easily compared using cosine similarity, as previously given by Equation 4.1. The CSD and EHD, which are parts of MPEG-7 content description, are employed based on the implementation by Bastan, Cam, Gudukbay, & Ulusoy (2010). The implementation provides other sophisticated visual features as alternatives of the currently used CSD and EHD.

5.2 Literature and Motivation

Here we present the literature of tagging related from the perspective of memory architectural model, which is closely related to personal memory organization and autobiographical memory.

5.2.1 Script and schema

In the early development of human knowledge structure representation, Schank (1982) and Schank & Abelson (1977) introduced the concept of *Script*, which is a kind of schema to represent knowledge of past events and experience. The way script represent events and past experience is defined to be goal directed and consist of sequence of either general or specific high level actions that can be perform. Figure 5.1 shows the restaurant script defined by Schank & Abelson (1977) to represent past experience of eating out in a restaurant in general. The script includes roles (i.e., cashier, waitress), and props (i.e., table, menu), however, not containing the detail of the food, decor, restaurant name, and amount to pay in the bill. The details can be inserted into relevant slots in the general script, and this is convenient considering the hierarchical structure of the script. Scripts explain the common observation in remembering routine, familiar, often-repeated events we seem to have a generic memory in which individual occasions, or episodes, have fused into a composite (Williams et al., 2008). The validity of scripts has been demonstrated by Bower et al. (1979) where, where they asked students to generate components actions that comprise an event, and list them in order of occurrence. They were asked about attending a lecture, visiting a doctor, shopping at a grocery store, eating at a fancy restaurant, and getting up in the morning. Figure 5.2 illustrates the lists for attending a lecture and visiting a doctor; where capital letters, italic, and lower case letters are mention by a large portion, a few portion, and the fewest portion of the students, respectively. We can see from the result that they ended up in a hierarchically structured with superordinate and subordinate goals, and the actions as well as the sequence in the scripts are agreed by the students. The scripts can also further divided into smaller scenes.

5.2.2 Schema copy plus Tag model

Regardless the positive outcome of the study by Bower et al. (1979), there seems to be discrepancies of the model with real life events experienced by human. The model assumes that the actions within the scripts must be familiar and general/routine activities.

Script:	Restaurant (the script header)
Roles:	Customer, waitress, chef, cashier
Goal:	To obtain food to eat
Subscript 1:	Entering
	move self into restaurant
	look at empty tables
	device where to sit
	move to table
	sit down
Subscript 2:	Ordering
	receive menu
	read menu
	decide what you want
	give order to waitress
Subscript 3:	Eating
	receive food
	ingest food
Subscript 4:	Exiting
	ask for check
	receive check
	give tip to waitress
	move self to cashier
	move self out of restaurant

Figure 5.1: An example of a restaurant script by Schank & Abelson (1977).

We need to consider the fact that there are unique events and one-off experiences, such as the day you graduate high school, or the time you won a lottery. Findings from Brewer & Treyens (1981) suggests that when memory was tested for objects in a room, schema-inconsistent objects were recalled better than schema-consistent objects. Interestingly, a study by Bower et al. (1979) with schema-inconsistent action turned out to be in a similar fashion with schema-inconsistent objects by Brewer & Treyens (1981). Nakamura et al. (1985) conducted an experiment in a form of lectures, where students attended a staged, 15-minutes lecture, which the lecturer performed several actions that varied in the relevancy to the lecture script. Figure 5.3 shows a list of relevant and irrelevant actions that was performed during the staged lecture. After the lecture, students were asked to recognized and identify the performed actions from a given list of actions. They found out that irrelevant actions were recognized better than the relevant ones, and the false alarm

Attending a lecture
ENTER ROOM
Look for friends
FIND SEAT
SIT DOWN
Settle belongings
Look at other students
Talk
Look at professor
LISTEN TO PROFESSOR
TAKE NOTES
CHECK TIME
Ask questions
Change position in seat
Daydream
Look at other students
Take more notes
Close notebook
Gather belongings
Stand up
Talk
LEAVE

Visiting a doctor
Enter office
CHECK IN WITH RECEPTIONIST
SIT DOWN
Wait
Look at other people
Name called
Follow nurse
Enter examination room
Undress
Sit on table
Talk to nurse
NURSE TESTS
Wait
Doctor enters
Doctor greets
Talk to doctor about problem
Doctor asks questions
DOCTOR EXAMINES
Get dressed
Get medicine
Make another appointment
LEAVE OFFICE

Figure 5.2: Script actions for the events of attending a lecture and visiting a doctor based on the study by Bower et al. (1979).

rate was three times higher for relevant actions. The result of the experiment were interpreted in a schema copy plus tag (SC+T) model proposed by Graesser (1981); Graesser, Gordon, & Sawyer (1979); Graesser & Nakamura (1982); Graesser, Woll, Kowalski, & Smith (1980); Smith & Graesser (1981), in order to explain the relevancy of the actions or objects related to the recalled memory occurred based on script-based passages. The term *tags* here are correspond to the irrelevant, unexpected, or deviant aspects of the event. They are also distinctive and memorable, and used for the retrieval of specific episodes. This notion of *tags* inspires us to incorporate the formal definition of context in our implementation of ERIS. Nevertheless, the schema copy plus tag itself is considered to be oversimplified to model the major phenomena within human memory, and a more dynamic model to represent past experience is needed.

Relevant actions
Sitting on a corner of a table
Pointing to information on the blackboard
Opening and closing a book
Moving an eraser to the blackboard
Handing a student a piece of paper

Irrelevant actions
Putting a piece of paper in a trash can
Scratching head
Wiping off glasses
Bending a coffee stirrer

Figure 5.3: Relevant and irrelevant actions considered in one of two lectures experiment in the study by Nakamura et al. (1985).

5.2.3 Motivation

After preceded by scripts and schema as well as the schema copy plus tag memory models, plus the thematic and temporal organization discussed in Section 3.5, Reiser, Black, & Abelson (1985) investigated the function of script-like knowledge structure in the process of organizing and recollecting past experience. They also compared the effectiveness of two different knowledge structure in terms of accessing personal memories. The two knowledge structures are called *activities* and *general actions*. Activities are script-like structures consist of knowledge in the form of sequence of actions in order to achieve a particular goal (e.g., eating in a restaurant, getting a medical check up in a hospital). On the other hand, general actions are higher level actions that can be performed in many different situation and context (e.g., buying tickets, reading a magazine). Reiser et al. (1985) also proposed *Activity Dominance Hypothesis*, to predict that activities would be better retrieval cues than general actions. This is due to accessing specific activities generates many inferences about food, decor, service that serves as further cues for retrieving specific details of memory. They asked subjects to recall specific personal experiences to fit a given activity cues such as *went out drinking*, and a general action such as *paid at the cash register*. The order of the two cues are varied to analyze the response time of the two orders, and they found out that the order with activity cues first and general action as the second cue yields faster response time. They concluded that there are two stages of personal memory retrieval to achieve optimal level of specificity: 1) establish a context, and 2) finding an index or tag that matches particular experience within that context. This model is called *context + index* model, which is similar to schema copy + tag model. However, several flaws of context + index model are pointed out by Conway & Bekerian (1987) and Barsalou (1988). Conway & Bekerian (1987) failed to replicated the findings of Reiser et al. (1985), and they suggested that retrieval of personal memory was based on lifetime periods (e.g., schooldays, college). Barsalou (1988) questioned the Activity Dominance Hypothesis and compared the effectiveness of a variety of different cues, such as activity cues (e.g., reading a book), participant cues (e.g., your mother), location cues (e.g., at home), and time cues (e.g., at noon). They found out that order of the cues is not an influencing factor in terms of retrieval time. They also suggested that goals (e.g., passing an exam or learning to drive) is a good cue to remembering personal memories.

The literature above inspires us to design similar mechanics for a robot to efficiently stores and retrieve personal memories, considering the interconnectivity between components. In particular, the feature of our proposed architecture incorporates the notion of context and tags for the familiarity mechanism.

5.3 Memory Processing Phase

Familiarity mechanism is closely related to how the personal memories of robot are processed. We consider four kinds of memory processing phase: memory formation, consolidation, revision, and recollection. Currently there are two memory types implemented, Semantic Memory (SM) and Episodic Memory (EM). As both types are parts of the Long-Term Memory, SM represents general facts of the world, and EM represents events-related information experienced by the robot (currently, it is limited to visual stimuli). To easily understand the principle of knowledge development process, we describe the four memory processing phase using illustrative scenarios.

5.3.1 Formation Phase

Let us assume that we present the robot a red ball. After the RGB image of the captured scene goes through the image processing module, the content of a SM item file is initially formed as a raw memory item to represent a raw knowledge, depicted in Figure 5.4a. The `OBJ_` identifier refers to a general object, as in the future work we may include other specific facts that can be represented, such as human face, a person, and animal. The header contains the encoding timestamp. Initially, the object comes with an empty array

```
OBJ_asd123.sm = {
  header=⟨timestamp⟩,
  label=asd123,
  tag=[],
  image=⟨BGR image of the ball⟩,
  color=⟨CSD⟩,
  texture=⟨EHD⟩
}
```

(a) An example of raw semantic memory item

```
SE_zxc456.em = {
  header=⟨timestamp⟩,
  label=zxc456,
  obj_count=1,
  seq_count=0,
  pos_minx=[5],
  pos_maxx=[10],
  pos_miny=[25],
  pos_maxy=[35],
  obj_name=[asd123],
  sequence=[]
}
```

(b) An example of raw episodic memory item

Figure 5.4: Structure of raw semantic and episodic memory as ROS message format.

of tag, assuming that the robot does not have the capability to autonomously associate the asd123 object with its current knowledge. The label for both raw SM and EM item uses a 6-digit random alphanumeric to prevent label duplication and acts as a temporary placeholder label. It also contains the BGR image of the detected object and the two features mentioned earlier.

Figure 5.4b depicts the formed raw EM item. The SE_ identifier refers to a *specific event*, which currently the only type of event implemented. The header and label serve the same purpose with the one in the SM item. Assuming that the scene zxc456 contains only a single object asd123, the obj_count has the value of 1, representing the object count at each scene. The seq_count and sequence corresponds to the number of robot movement sequence from the Procedural Memory (PM) during the scene and the array of the name of the sequences, which currently serves no purpose yet as the PM details are not yet implemented. Since the scene is supposed to contains objects, the 2D plane position of each object with respect to the Field of View are available in the form of array, with the variable of pos_minx, pos_maxx, pos_miny, and pos_maxy. Finally, the obj_name corresponds to the object label detected at the scene. This information is interconnected with knowledge within the semantic memory, because in order to know what object is detected at a particular scene or event, the knowledge regarding that particular object must be exist, and semantic memory is the proper place to store these kind of knowledge.

Table 5.1: Summary of cues of memory item as ROS message

Category	Identifier	Remarks
Filename prefix	OBJ_	indicating objects
Filename prefix	SE_	indicating specific event
File extension	.sm	semantic memory extension
File extension	.em	episodic memory extension
General-purpose cue	header	ROS header msg
General-purpose cue	label	indicating label of a particular memory item
Context-dependent cue	tag	storing tags from interactions
Placeholder value	$\Big\langle \dots \Big\rangle$	asdasd
1D Array value	$[\ v_1, \dots, v_n\ ]$	given n as the array size and $1 < m < n$, v_m is the element value of the array.

5.3.2 Consolidation Phase

To distinguish the raw and revised knowledge, a database is used to track all the known information based on the interaction with human. Only revised knowledge will be listed in the database. Initially, when there were no revised knowledge possessed by the robot, the database is empty. After a memory is formed, it is consolidated to the Long-Term Memory Storage as its filename (e.g., OBJ_asd123.sm or SE_zxc456.em). Depending on the type of the knowledge, the consolidation process consists of:

1. storing the formed memory into the hard-drive as the corresponding memory files (for both raw and revised knowledge), and
2. recording and updating the memory item information within the database (only for revised knowledge).

For a raw memory item processing (representing raw knowledge), no database operation is performed, because there were no information provided for interaction by human during the capture time of the scene.

5.3.3 Revision Phase

Up to this point, the memories formed in Figure 5.4 are called *raw knowledge*, as discussed in Section 4.1. When the robot detects a previously seen object, it recognizes the object as the corresponding memory exists in the long-term memory. With this mechanism, the robot is able to develop its own knowledge even without human interference.

However, we expect the robot to interact with human at a certain period of time. When a human verbally ask if the robot knows what is the object called, it will respond with nothing regarding the object label, because the robot does not have the information about the object label that the human supposed to understand. It will need to interact with human to provide meaningful information to the known knowledge for future occurrence of human-robot interaction.

Then, notion of *knowledge revision* plays a role here. The knowledge revision process involves the notion of **tagging** of memory items. In order to represent different human-human and human-environment interactions, it is believed that neural equivalents of placeholder labels are used to refer to them (Manis & Meltzer, 1978). A tag is a symbolic reference used to categorize factual knowledge in the form of a Semantic Memory item (i.e., an object or a related concept). In Semantic Memory, each memory item s_i is characterized by a label λ_i, a set of *tags* $\Psi_i = \{\psi_{i,1}, \ldots, \psi_{i,n}\}$, which allow it to be referenced by a unique label and multiple associated tags, and physical properties of the objects, such as visual features $\Phi_i = \{\phi_{i,1}, \ldots, \phi_{i,n}\}$, such that $s_i = (\lambda_i, \Psi_i, \Phi_i)$, $i = 1, \ldots, S$. This correspondence entails an associative property: memory items in Semantic Memory are directly connected to relevant tags and each tag to the associated memory item. Figure 5.5 depicts the sequential process of knowledge revision including the tagging process. During a first initialization phase, a human requests a specific memory item to be revised. Then, if such memory item exists, it is associated with a label and a set of tags. The revised memory item is validated and consolidated in long-term memory.

Continuing our example, human symbolically tell the robot that the object that it saw was a *redball* and it is associated to the tag *round*, and *plastic*. After revision, both raw SM and EM item in Figure 5.4 become revised memory items (representing revised knowledge) as in Figure 5.6.

Revising a memory item involves a consolidation process including the database update operation to make sure that the new information within that memory item is properly updated. Therefore, after the revision example above, the database file _DB.ffi is updated and contains of the following information:

```
_DB.ffi={
  known_objects=[redball]
  known_tag=[round, plastic]
}
```

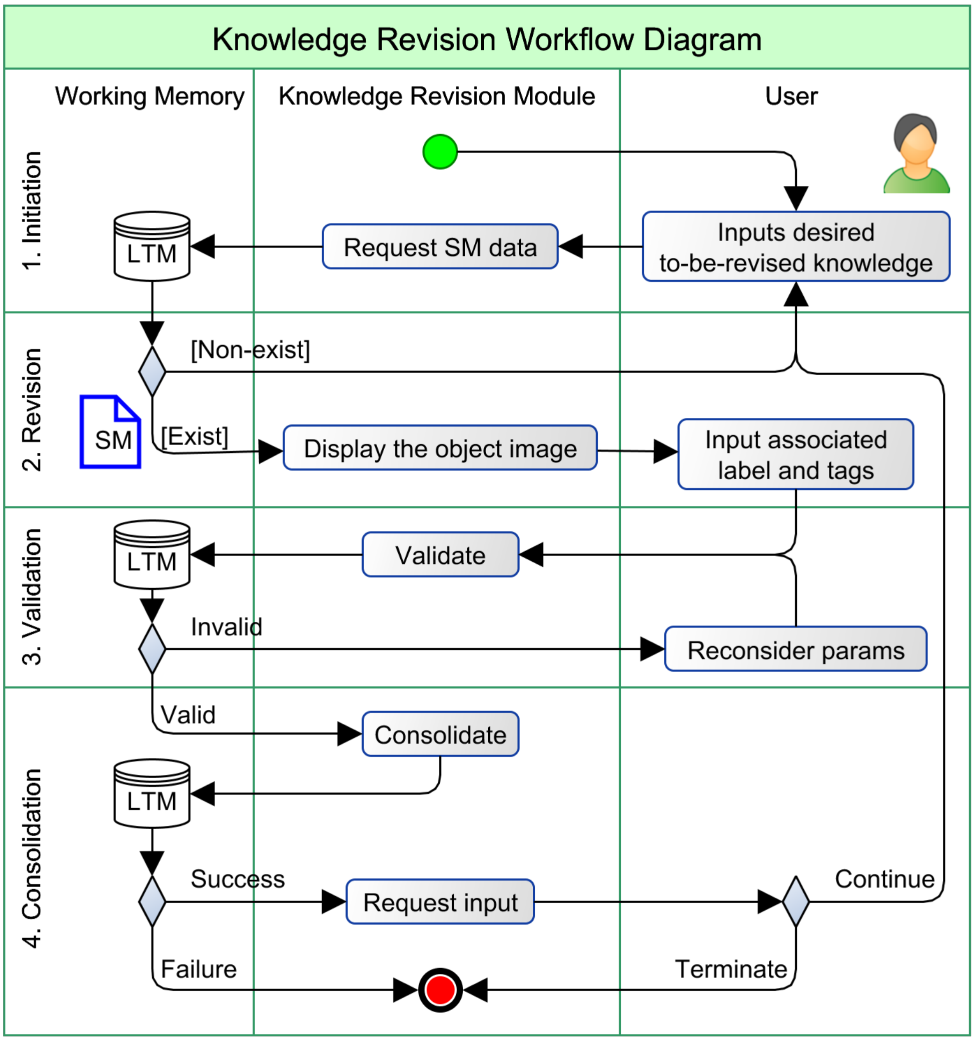


Figure 5.5: Tagging in a human-assisted knowledge revision process.

5.3.4 Recollection Phase

Next, the memory recollection involves checking the Long-Term Memory storage and retrieve the desired information. This process occurs in the **weak** sense of *conscious* and *unconscious* recollection. Unconscious recollection means that when the robot is about to determine whether an object is familiar, it should recall its current knowledge regarding that object. Meanwhile, conscious recollection occurs during human-robot interaction, where a human demands a specific information regarding a particular knowledge of a certain event or fact.

Two kinds of familiarity measures are considered:

1. object familiarity, which is related only to Semantic Memory; and

```
OBJ_redball.sm = {
  header=⟨timestamp⟩,
  label=redball,
  tag=[round, plastic],
  image=⟨BGR image of the ball⟩,
  color=⟨CSD⟩,
  texture=⟨EHD⟩
}
```

(a) An example of revised semantic memory item

```
SE_zxc456.em = {
  header=⟨timestamp⟩,
  label=zxc456,
  obj_count=1,
  seq_count=0,
  pos_minx=[5],
  pos_maxx=[10],
  pos_miny=[25],
  pos_maxy=[35],
  obj_name=[redball],
  sequence=[]
}
```

(b) An example of revised episodic memory item

Figure 5.6: Revised semantic and episodic memory as ROS message format.

2. scene familiarity, related to both Episodic Memory and Semantic Memory.

During the human-robot interaction process, inquiries can be made to the robot by specifying the context in which to frame the response. As of Definition 11, a **context item** γ_j is characterized by a **retrieval cue** c_j, a **value** for that cue v_j, and a set of relevant **tags** Ψ_j, such that $\gamma_j = \{c_j, v_j, \Psi_j | 1 < j < X\}$. A context item refers to lexical information, which consists of a single cue, its value, and a set of tags, which are expected to match the whole or part of the set of tags provided during the knowledge revision process. Then, a **context** ξ is modeled as **a set of** X **context items** γ_ξ, such that $\xi = \{\gamma_{\xi,1}, \ldots, \gamma_{\xi,X}\}$.

In Episodic Memory, the memory item is called an **episode**, formalized as $\hat{\sigma} \in EM$, which is a **digest of a scene**. A scene σ is a captured visual stimuli, resulting from the changes of visually detected input, which indicates the occurrence of an event at a particular time. In short, a **scene** is an **event marker**. Anything occurs between two distinct scenes is defined as an **event**. Scenes that have been captured are formed into episodes, and stored in the LTM. An event η is associated with, and occurred over a period of time, which marked from two distinct scenes correspond to the beginning and the end of an event. It consists of multiple, timestamp-ordered episodes during the period of that event, defined as $\eta = \{\hat{\sigma}_{\eta,1}, \ldots, \hat{\sigma}_{\eta,s}\}$, given s is the number of episodes for that particular event.

Figure 5.7 depicts the representation of events and objects, as well as the relations between them. Any detected objects within a scene are represented as Semantic Memory

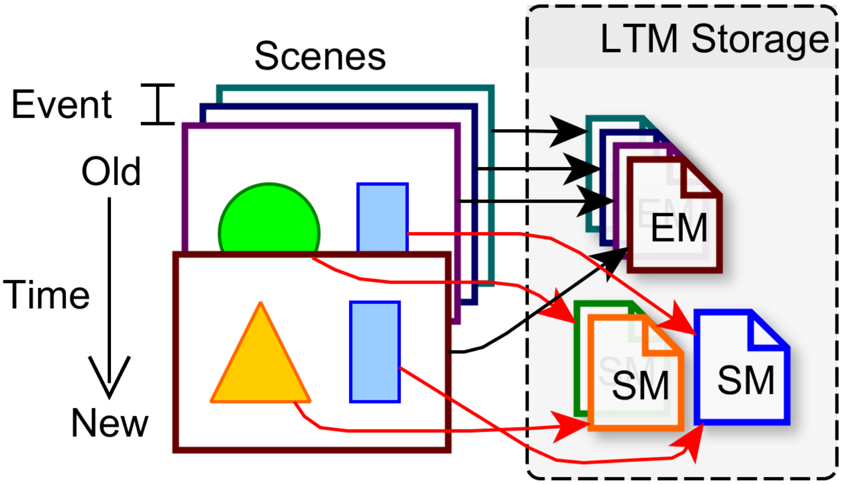


Figure 5.7: Representation of objects and events, and their relations.

items, and familiar objects detected in a scene with respect to the past experience refer to the same Semantic Memory item (e.g., the blue box in Figure 5.7 in the newer scene refers to the previously consolidated Semantic Memory item from the older scene). As changes occur between scenes, unfamiliar objects are consolidated as raw Semantic Memory items and the corresponding scene as an Episodic Memory item (episode).

Up to this point, one cycle of the knowledge development process is complete. As long as ERIS is activated within the robot, the cycle keeps continue and the robot keeps gathering knowledge into memories and personal experience from the received stimuli. In some occasions, the cycle could be either [recollection-formation-consolidation] when no interaction with human is considered, or [recollection-revision-formation-consolidation] when the interactions are considered. In principal, the cycle [recollection-formation-consolidation] emphasizes the capability of **progressive knowledge development** without human intervention or assistance.

5.4 Memory processing limitations and assumptions

We consider **two equal scenes in a different timestamp as mutually exclusive**, hence if a familiar scene is detected, it will not be consolidated since we have the first occurrence of the scene as the episode from the past experience. The implicit assumption that we take into account is that every object contained in a scene must have the corresponding Semantic Memory item. In other words, when a robot recalls a particular scene of its personal experience, it must know what objects exist in that scene by having their Semantic Memory items in the Long-Term Memory storage, regardless of being raw or revised knowledge.

For the object familiarity, we currently consider color familiarity using the implementation of Color Structure Descriptor (CSD), and Edge Histogram Descriptor (EHD) for the texture and shape familiarity. Both descriptors are part of the MPEG-7 content description standard (Manjunath et al., 2002). We also consider that the objects used in the experiments have a distinctive colors that are contrast enough to be detected in the workspace, in this case a flat, white surface table. The text-based communication is currently implemented as simplified grammar with no sophisticated natural language processing interface, so the input set from the human will be directly processed based on the formal definitions without involving a sentence parsing process, and the HRI interface module only gives the readers several examples out of many possible questions to ask the robot integrated with ERIS, which demonstrates the versatility of verbal interaction with ERIS developmental structure. We assume a static field of view of the main input camera, which affect the scene familiarity. Subsequently, this bring us to the limitation that a scene is considered familiar when detected objects are determined to be familiar and their spatial differences with respect to the field of view are within a predetermined threshold value. Occlusion between detected objects and forgetting mechanism are not considered at the current development stage.

As discussed in the previous subsection, in principle, the cycle [recollection-formation-consolidation] emphasizes the progressive knowledge development with no human assistance. However, in our current development stage, the captured scenes are segmented manually, meaning that there is human assistance in capturing the scenes. We consider this as a segmentation problem, which is directly unrelated to the proposed architectural design, and will be resolved in the future development stage.

5.5 Different Cases of Memory Recollection

We consider three fundamental cases where memory recollection is performed, which are:

1. perception of visual stimuli only;
2. lexical context only; and
3. perception of visual stimuli complimented with lexical context.

Those three cases involves **object recollection** and **scene recollection**. From this distinction, we shall see how inputs within each case affect memory recollection.

During the object recollection for Case 1, this process deals only with SM item. Given an image referring to an object, a human inquire the robot regarding the familiarity status of the detected object. In case of object recollection within a scene or during an occurrence of an event, please remember the implicit assumption we consider in Section 5.4, that the robot must know the details of objects information contained in a scene during a memory recollection process, by having the corresponding SM item in the Long-Term Memory storage. Therefore, the robot will respond with nothing when a query related to a novel object is inquired. The scene recollection deals with both EM and SM, and given a scene image, a human inquire the robot regarding the familiarity status of that particular scene.

The given input set of lexical context in case 2 and case 3 is defined according to the context item notation in Definition 11. For case 2, given the input context, a human then inquire the robot about the familiarity status of the context, which may involve either episodes or SM items.

Case 3 is simply a combination of case 1 and case 2. Similarly, this can refer either to object or scene recollection. For the object recollection, given an object image and contextual information, a human inquire the robot about the familiarity status of an object that conceptually looks like the input image, but possess different characteristics lexically described as the context. For the scene recollection, given a scene image and contextual information, a human inquire the robot about the familiarity status of a particular scene with the characteristics described in the given context.

This is an important feature of ERIS, as **it allows the robot to retrieve different knowledge based on various, multiple contexts from its current knowledge and/or the given contextual information**. Memory recollection applies to *conscious* and *unconscious* process. We refer to it as a *conscious* process when it occurs as part of a human-robot interaction process (i.e., case 1, 2 and 3), and as an *unconscious* process (passively perceiving the environment) during knowledge acquisition (i.e., case 1 only).

Next, a more advance memory recollection procedure is also applies to exploit the past experience of the robot, as a part of the human-robot interaction conducted in our experiment. Such procedure allows human to ask the robot regarding the facts within possible past events, without limitations only to familiarity. Some examples of the applicable questions to the robot are *"Have you been presented with a purple box before a green ball?"*, *"Did you move the blue box to the right hand side?"* or even *"How many objects* *do you know that are related to a box?”.* This represents **the ability to recall a general knowledge regarding all the knowledge** that the robot has experienced.

Chapter 6

Experiment 1: Manifestation of Robot Personal Experience

> Character is higher than intellect. A great soul will be strong to live as well as think.
>
> Ralph Waldo Emerson

6.1 The Properties of Events

ERIS allows a robot to gain knowledge as time progresses, and the concept of events allows further potential of exhibiting past experiences flexibly based on the contents within episodes. As we have discussed in the explanation for Definition 9 in Section 4.3, the concept of event is characterized by three interesting properties, which are flexible, subjective, and personal. **Flexible** means there are no strict definitions of atomic or compound events (on the contrary to what has been postulated in the work by Stachowicz & Kruijff (2012)). Due to this property, any two events can be distinct, overlap in terms of the events timestamps, or one can include the other, which also complies with the whole set of relationships between intervals defined by Allen (1983). **Subjective** means that an event may be interpreted in multiple ways, depending on the presence of objects in the scene during the occurrence of the event. An event is subjective when, for instance, event A = {episode 1, episode 2, episode 3} where all episodes in event A contains a blue box on the left hand side, and a red ball on the right hand side. We consider an episode is a collection of symbols representing specific properties of the scene. Therefore, during human-robot interaction, if a human inquire the robot the familiarity status of an event with a blue box on the left hand side, the robot will respond with event A as the answer. Assuming that the robot only experience one event (event A) so far, if a human inquire the familiarity status of an event with a red ball, it will also respond with event A. This illustrates that the same event may be interpreted in different ways depending on the contents of the episodes and the request/desired situation. Finally, **personal** means events between two identical robots, each exposed with similar visual stimuli in the same architecture might be different due to the difference in the previously gained knowledge. Interestingly, although addressed from the event perspective, this personal property seems to share common idea with **The Principle of Subjectivity** addressed by Stoytchev (2009), which emphasizes the interaction history during human-robot interaction.

In our architecture, this phenomenon is exhibited fundamentally by processing non-familiar scenes and consolidating the related memory items. Figure 5.7 illustrates the phenomenon, as an episode is representing a scene, which is a captured visual stimuli of the environment. Consolidating the episode means representing a specific snapshot of visual experience in a particular period of time. Recognition of familiar scenes or objects is analogous to remembering certain past events, and their corresponding memory item recollection processes are subsequently performed.

6.2 Motivation for a small scale experiment

Before going directly into the long-term experiment, we shall test our proposed system in a small scale experiment, which consists of a small number of scenes experienced by the robot. Ideally, in the long term experiment, the robot should be able to progressively gaining knowledge and at the same time handle the interaction with human and the environment for days, months or even longer. Even though the robot is turned off for a certain period of time, the knowledge (in the form of nonvolatile memories) should not be influenced during the deactivation period. After the robot is turned back on, the robot should retain the memories and continue the knowledge development from that point. In this early development stage of ERIS, it is a good chance (and easier) to explain the fundamental phenomena of the progressive knowledge development that are exhibited by the robot using a small amount of scenes.

6.3 Experimental Design and Evaluation

6.3.1 Experimental Scenario

In this section, we describe the experimental scenario designed to emphasize the features introduced in the previous chapters. The target robot platform is Baxter, a dual-arm manipulator from Rethink Robotics. The robot workspace is constituted by a flat, white surface table with a dimension of $0.5m \times 2m \times 0.7m$, where we located objects with various shapes and colors. This is also to make sure that the color of the objects are contrast to the white surface table. The objects have similar dimension, approximately $0.1m \times 0.1m \times 0.1m$, which are easily graspable using Baxter's gripper. The camera integrated in the Baxter's left hand is directed towards the objects, and kept at a fixed configuration, as shown in Figure 6.1. The experiment is made up of two phases:

1. progressive knowledge acquisition; and
2. memory retrieval through interaction.

Initially, Baxter's LTM storage is empty and it has no memory items, nor it has any notion of familiarity. During the first phase, Baxter performs pick and place movements in sequence, moving specific objects using the gripper of the right arm. This phase represents changes in the environment, which also includes the detection of novel objects and the removal of existing objects within the robot field of view. A scene is manually segmented whenever a pick and place action is completed. Examples of scenes captured in this phase are depicted in Figure 6.2. Then, during the second phase, a human can ask Baxter about the familiarity status related to specific input at any given time, according to any of three cases introduced in Section 5.5. Technically, phase 2 can be initiated at any time during phase 1, because an epigenetic robot should be able to perform phase 1 without any human interference and human may interact with robot at any given time. However, here we clearly separate both phases since we would like to consider all the experience obtained by the robots before the human-robot interaction. Occluding configurations among objects and forgetting mechanisms are not considered.

After the first phase is complete, knowledge revision can be performed according to the procedure depicted in Figure 5.5, and known objects are associated with a label and multiple tags. The relationship among tags, SM and EM data is shown in Figure 6.3. During the first scene segmentation as depicted in Figure 6.2a, two unidentified objects labels with their position with respect to the robot field of view are consolidated as episode

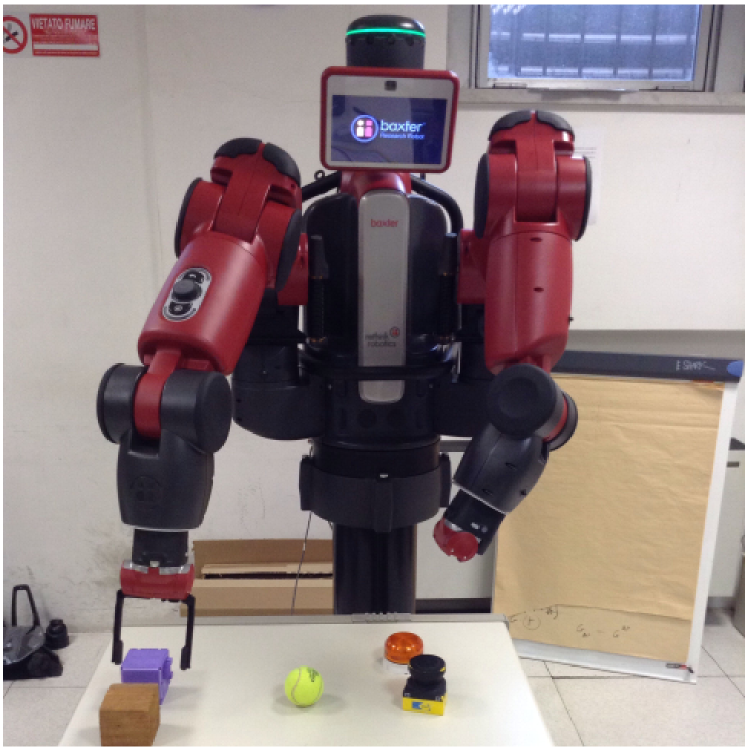
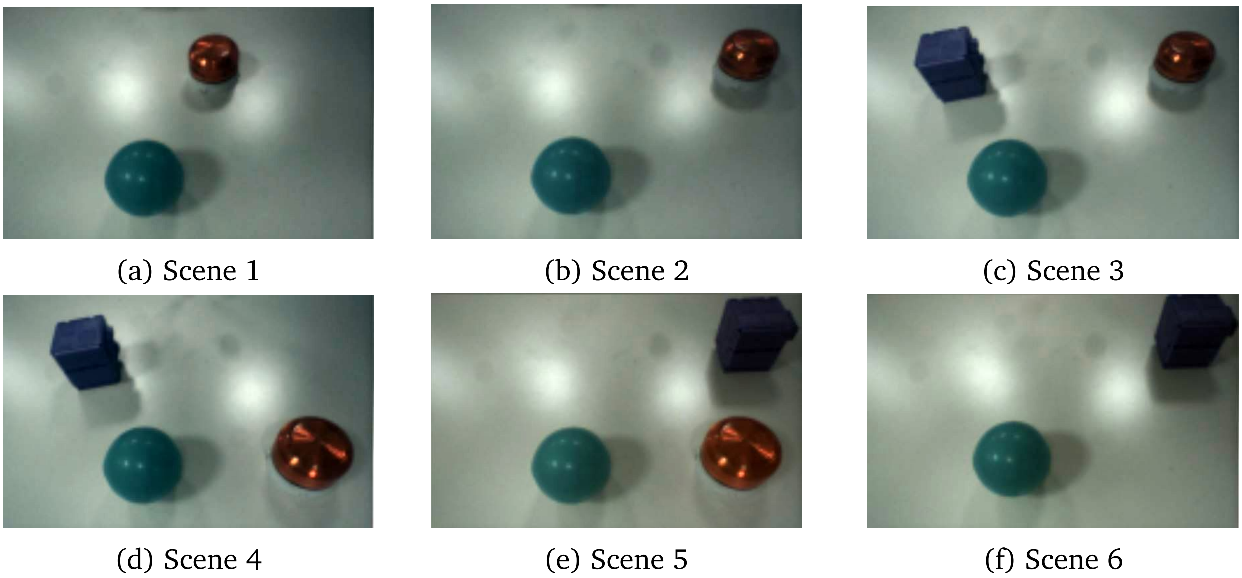


Figure 6.1: Baxter witnessing the event

1. As time progresses, since no novel objects are detected and all the detected objects are familiar to Baxter with different location, no new SM data is created and only episode 2 is consolidated. Since Baxter has the information of all familiar objects, the objects information in episode 2 are referring to the existing SM data. So far, Baxter has two episodes (episode 1 and episode 2) and two unlabeled SM data with color and texture representation and no associated tags. As a novel object is presented to Baxter, a new scene is segmented and consolidated. It is labeled as episode 3, which contains three objects with their corresponding information, which one of them is a novel object. In addition, a new SM data correspond to the novel object is created. Finally, at the end of the consolidation of episode 3, Baxter has three episodes and three SM data. At this point, a knowledge revision process is commenced, where a human assigns tags that are associated with the objects. After revision, the object labeled as orangelamp is assigned the tags orange, lamp, and round. Since EM and SM are interconnected, the label of the orangelamp in episode 1, 2, and 3 are automatically updated from the less meaningful, random alphanumeric placeholder label.

The tags are only added in the SM because it is independent with respect to past events. To make things more interesting, some common tags (e.g., round, toys) are as-



(a) Scene 1 (b) Scene 2 (c) Scene 3

(d) Scene 4 (e) Scene 5 (f) Scene 6

Figure 6.2: Scene configuration dataset

signed to the known objects. The object tealball is assigned the tags round, teal, and toys, and cube is assigned the tags toys, purple, and cube. The same procedure applies to the rest of the scenes captured, as shown in Figure 6.2. Please note that other details within EM data such as object count, timestamp, scene label are omitted for clarity. Now, let us see the three interaction cases in a bit more detail.

Case 1. In order to verify the robot capabilities to exhibit object familiarity, we present three objects to Baxter one at a time: every time, the robot provides the familiarity status related to the presented object. For the scene familiarity, we arrange the objects on the table to resemble specific experienced scenes and also set a completely new workspace arrangements to see whether the witnessed scenes are familiar to Baxter.

Case 2. We query Baxter about the familiarity status of contextual information via a command-line interface. For the object familiarity, this will only correlate with known SM data. In this case, we expect a Boolean response for the familiarity status, with the addition of objects that are similar to the context. For the scene familiarity, the input is contextual information as formally defined in Definition 11, and will correlate with either SM, EM, or even both, depending on the context.

Case 3. For the object familiarity, several objects are presented one at a time, with additional contextual information regarding the object characteristics that we would like Baxter to remember. For the scene familiarity, the workspace is rearranged in various configurations and a scene is captured for each configuration as the input for Baxter with additional contextual information.

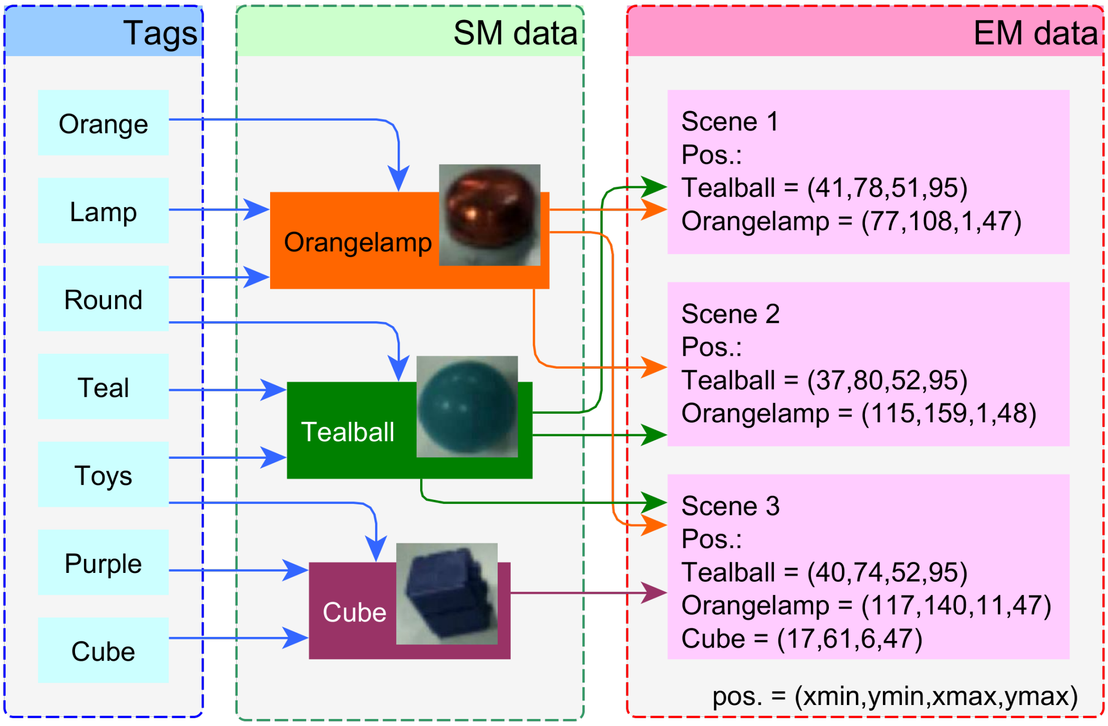


Figure 6.3: Object-tag associations after revision for scene 1 to 3

6.3.2 Results

The results for case 1 are shown in Figure 6.4. Six objects are presented to Baxter one at a time for the object familiarity, and the workspace is rearranged as the depicted configurations. Table 6.1 shows the essential results for case 2. Aside from Boolean responses for the context in Table 6.1a, the interface may also provide details about objects that support a Boolean response or similar to the desired input. Figure 6.5 presents both object and scene familiarity results for case 3. For the object familiarity, since the criterion for similarity matching is based solely on color, only objects that were previously presented to Baxter are tested, so the influence of additional contextual information given aside from the presented object can be easily analyzed.

6.3.3 Discussion

Let us discuss the results for case 1. For the object familiarity test, three objects are identified to be familiar and other three to be unfamiliar due to the absence of the corresponding SM data. As pointed in Section 5.4, object familiarity matching is currently based on color and texture. The future work will be focused on using a dynamic robot

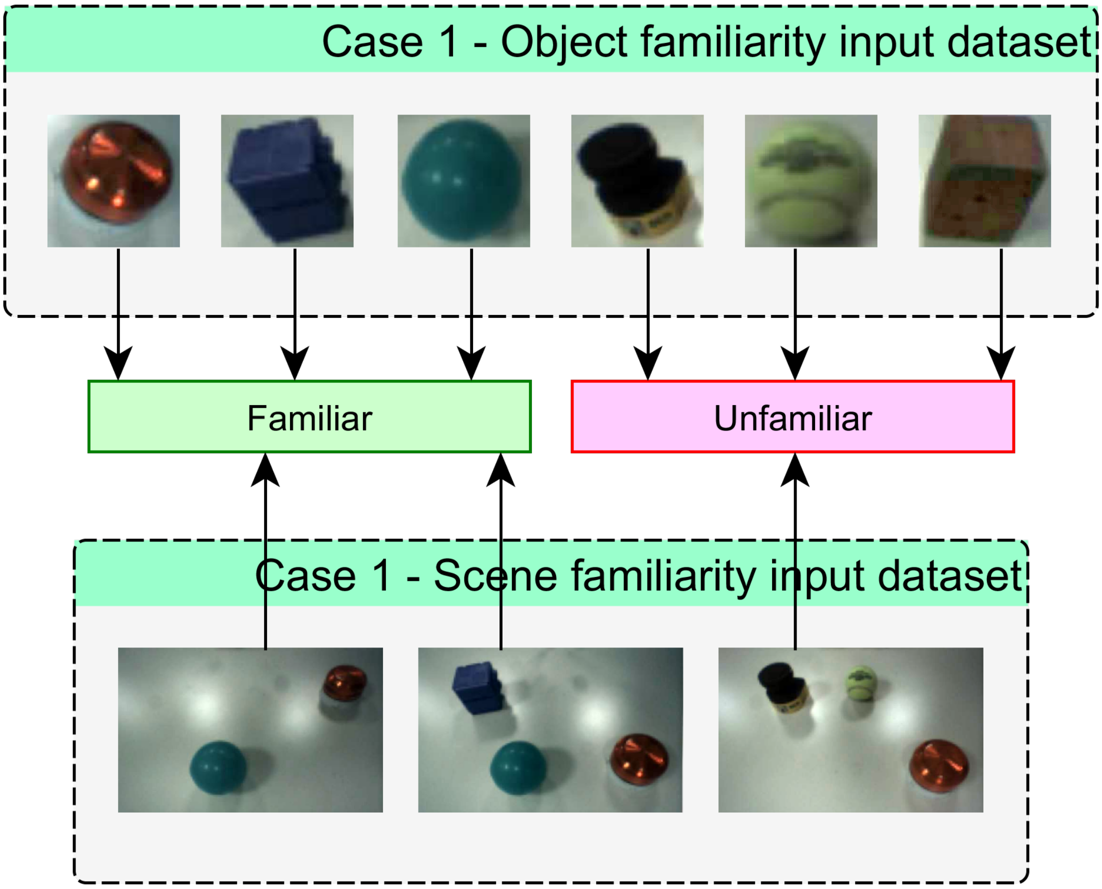

Case 1 - Object familiarity input dataset
Familiar
Unfamiliar
Case 1 - Scene familiarity input dataset

Figure 6.4: Experimental results for case 1

field of view to achieve a more sophisticated matching.

In the scene familiarity, three different arrangements of the workspace were set up and identified as familiar for two arrangements and unfamiliar for one. So far, Baxter is able to recognize familiar situations based on a single input, either a single object or a particular workspace arrangement.

With a different approach, the input for case 2 is only constituted by a lexical context instead of a direct visual stimuli. This affects the overall processing of memory recollection. For the object familiarity, since the object label is desired to be identified, as in Definition 11, a context item γ_j is characterized by a **retrieval cue** c_j, a **value** for that cue v_j, and a set of relevant **tags** Ψ_j, such that $\gamma_j = \{c_j, v_j, \Psi_j | 1 < j < X\}$. In this case, the tags Ψ_j are listed in Table 6.1a. The context item of Tennis ball is identified to be unfamiliar. However, Baxter is able to pinpoint the object that is related to ball, by checking all the object label and associated tags, and therefore a Tealball is suggested as the closest familiar object with Tennis ball. The context item Green box, formalized as

Table 6.1: Experimental results for case 2

(a) Object familiarity with suggestions

Tags	Familiar	Suggestion
Tennis ball	No	Tealball
Green box	No	-
Round objects	Yes	Orangelamp, Tealball

(b) Scene familiarity with corresponding scenes

Contextual information	Familiar Scene
a ball at the leftmost	Yes, scene 1,2,5,6
3 objects with a cube at the rightmost	Yes, scene 5
yellow toys	No

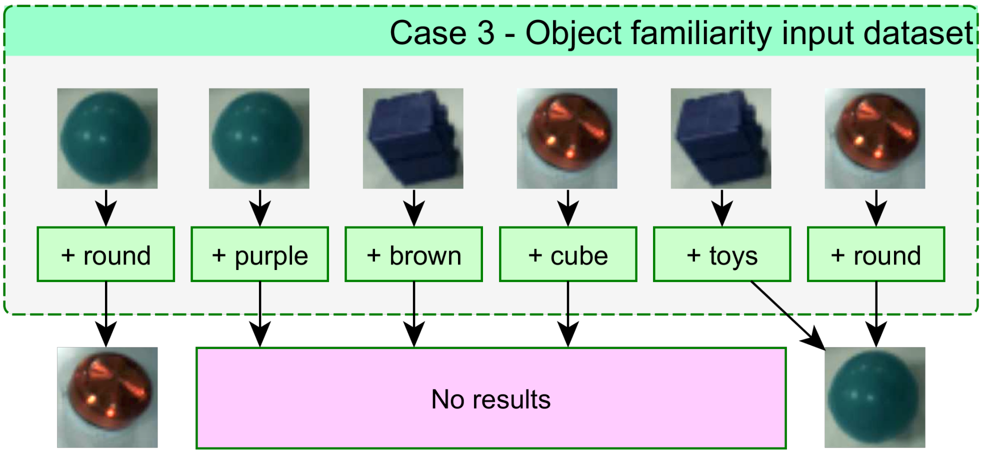
$\{c, v, \Psi\} = \{\text{object,label, green box}\}$, is identified as unfamiliar with no suggestion given. Similarly, the context item round objects, allows Baxter to give multiple suggestions aside the familiarity status of the input query. Here, Baxter is familiar with round objects, and all known round objects are listed, such as Orangelamp and Tealball.

Moving on to scene familiarity in case 2, a number of simple mapping indicators were implemented to achieve a more variety of inputs, such as leftmost and rightmost. As the position of the object with respect to the field of view is stored within the captured scene, given n as an object, δ as a scene, and pos_x as the x value of position with respect to the fixed field of view, leftmost and rightmost are simply defined as $\forall n \in \delta, \arg\min_{n \in \delta} pos_x(n)$ and $\forall n \in \delta, \arg\max_{n \in \delta} pos_x(n)$, respectively. With the first context given to Baxter, the whole question can be interpreted as *Are you familiar with a scene with a ball at the leftmost?*, and yields the result of scene 1, 2, 5, and 6. Considering the definition of *event*, the architecture may interpret this result one step further by inferring that an *event A* occurs between scene 1 and 2, and an *event B* occurs between scene 5 and 6. Hence, event A and B are the events occurred when a ball is located at the leftmost. Since the definition of event A and B depends on the experienced visual stimuli, other robots with the same architecture may have a different experience, hence resulting in different definitions of events. The second question is the familiarity of *a scene where three objects were presented with a cube at the right most* and yields familiar result with scene 5 as the only identified familiar scene. Since only one scene yielded as the interaction result, we may wonder how an *event* of three objects with a cube at the rightmost is defined. Since there are no multiple scenes matching this criterion, that particular event cannot be defined and considered non-exist. This is because an **event** is **dynamic**, meaning that changes within the field of view are to be expected during the occurrence of an event. On the contrary, a **scene** is **static**. Therefore, we can consider that at a certain point of time (i.e., at scene 5), a workspace configuration matched the criterion. As in the third context, we can also give a simple context yellow toys, and this yields unfamiliar results for a scene with yellow toys. From these results so far, Baxter shows the capability of recalling its personal past events given the contextual information, and the results clarify the properties of events: flexible and subjective. The personal property of event will become apparent later in the experiment discussed in the next chapter.

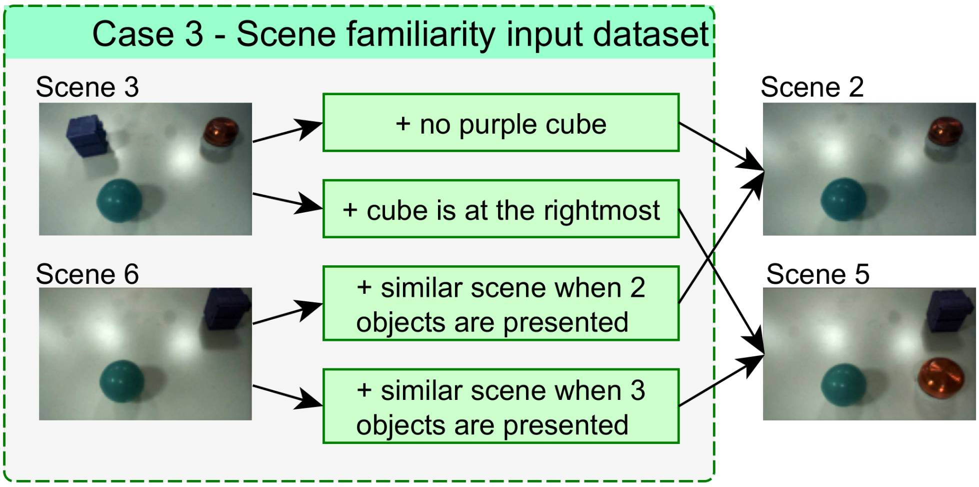
Case 3 involves visual stimuli and contextual information as the input. For the object familiarity, presented objects and the given context are shown in Figure 6.5a. Three previously presented objects are paired with various contexts to see how familiarity is affected by the context. For instance, the orangelamp is presented with an additional context item of round and yields the tealball as the object that is also associated with round category; the purple cube with the context item toys and tealball with the context item round. Other examples of unfamiliar results yielded when the presented objects do not match with the given context.

Figure 6.5b shows two distinct workspace arrangements monitored by Baxter, and additional contexts are given to check the familiarity status. Similarly, only familiar workspace arrangements are shown, so the influence of context can be easily seen. Therefore, an apparently unfamiliar arrangement with any context yields an unfamiliar result. We can see that the arrangement of scene 3 and scene 6 pairs with different contexts yield different familiarity results. This result demonstrates the influence of context through the personal scene recollection process.

This concludes our analysis about the manifestation of Baxter's personal experience through a descriptive experimental scenario, as well as the significance of contextual information during human-robot interaction.



(a) Object familiarity result


Case 3 - Scene familiarity input dataset
Scene 3
Scene 6
+ no purple cube
+ cube is at the rightmost
+ similar scene when 2 objects are presented
+ similar scene when 3 objects are presented
Scene 2
Scene 5

(b) Scene familiarity result

Figure 6.5: Experimental result for case 3

Chapter 7

Experiment 2: Exploitation of The Principle of Subjectivity

> Knowledge is true opinion.
>
> Plato

As discussed by Stoytchev (2009), the Principle of Subjectivity is extremely relevant in an epigenetic robot when we consider the robot interaction with human or the environment. Here, we illustrate our concrete example of two identical robot architecture that may have different knowledge representation given similar visual stimuli of the environment, and different interaction history with human through contextual information. First, using one unit of robot, two identical instances of ERIS are executed independently, within the same workstation. For conciseness, we will refer them as Robot A and Robot B. Both robots are subject to be given a similar visual stimuli and experience different interaction with human via knowledge revision process. Similar to experiment 1 discussed in the previous chapter, this experiment is also made up of two phases:

1. progressive knowledge acquisition;
2. memory retrieval through interaction.

Phase 1 is the process when the robots experiencing each of their given stimuli one after another, which occurs without any human supervision. Each of the memory representations will be created autonomously, as discussed previously. As soon as phase 1 ended with the processing of the last detected scene, we proceed to the phase 2, where a human may ask questions to Baxter through the HRI module about the past experience.

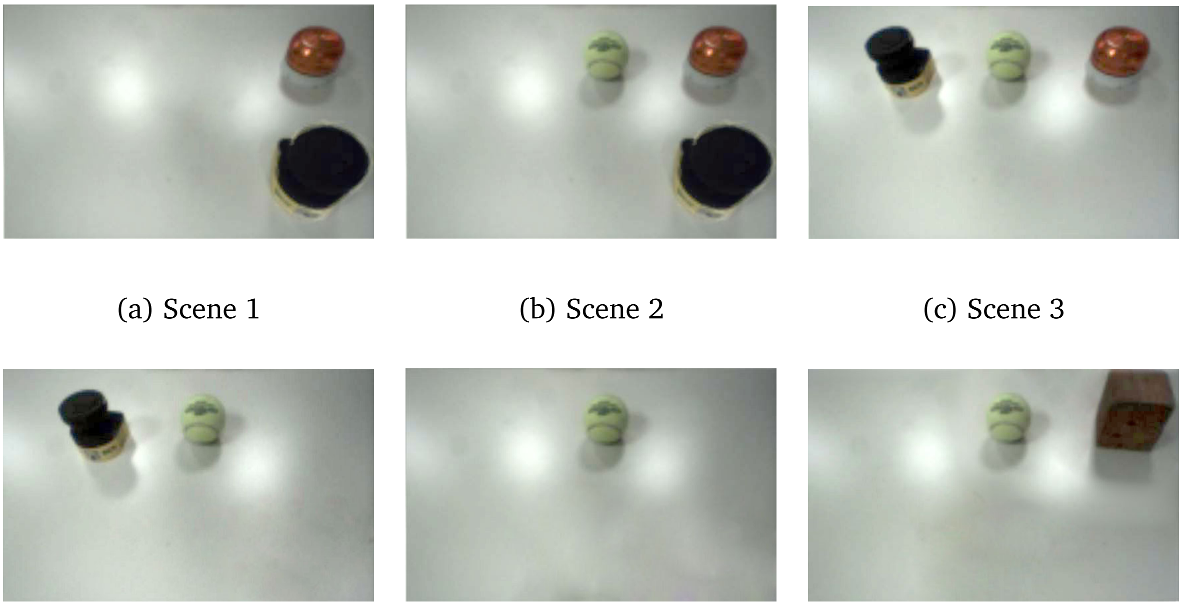
Technically, phase 2 can be initiated at any time during phase 1, because an epigenetic robot should be able to perform phase 1 without any human interference and human may interact with robot at any given time. However, here we clearly separate both phases since we would like to consider all the experience obtained by the robots before the human-robot interaction. At the end of the experiment, we expect both robots to comply with The Principle of Subjectivity, which is to have different knowledge representation and exhibit dynamic interaction with human, given the fact that both robots have different interaction history.

7.1 Experimental Design

To see the principle of subjectivity in a better perspective, instead of comparing ERIS with other epigenetic architectures, we do a side-by-side comparison of both Robot A and Robot B (represented by one unit of Baxter Research Robot), with identical architecture and the same initial condition (no prior knowledge and past events), and given different stimuli based on the case as previously elaborated.

This experiment serves as an extension of the experiment 1 discussed in the previous chapter. Similar with the previous experiment, Baxter robot serves as the robot platform, and the robot workspace is constituted by a table where we located objects with various shapes and colors. The camera integrated in the Baxter's left hand acts as the main visual capture device, which directed towards the objects and kept at a fixed configuration, as shown in Figure 6.1. This time, unlike the previous experiment, using the same robot, two different sets of workspace configurations are provided to the two identical architectures which carried out independently, which we have decided to refer them as Robot A and Robot B as previously mentioned.

Initially, for both Robots, Baxter's Long-Term Memory storage is empty and it has no memory items, nor it has any notion of familiarity. Phase 1 is conducted as either (a) a human present an object into or remove one from Baxter's Field of View; or (b) Baxter performs pick and place movements in sequence, moving specific objects using the gripper of the right arm. This phase represents changes in the environment, which also includes the detection of novel objects and the removal of existing objects within the robot Field of View. It is noteworthy that no exploration strategy is considered. A scene is manually segmented whenever a pick and place action is completed by a human or Baxter. The visual stimuli received by Robot A and Robot B are listed in Figure 7.1 and Figure 7.2, respectively. The second phase covers the human-robot interaction procedures via the



(a) Scene 1

(b) Scene 2

(c) Scene 3

(d) Scene 4

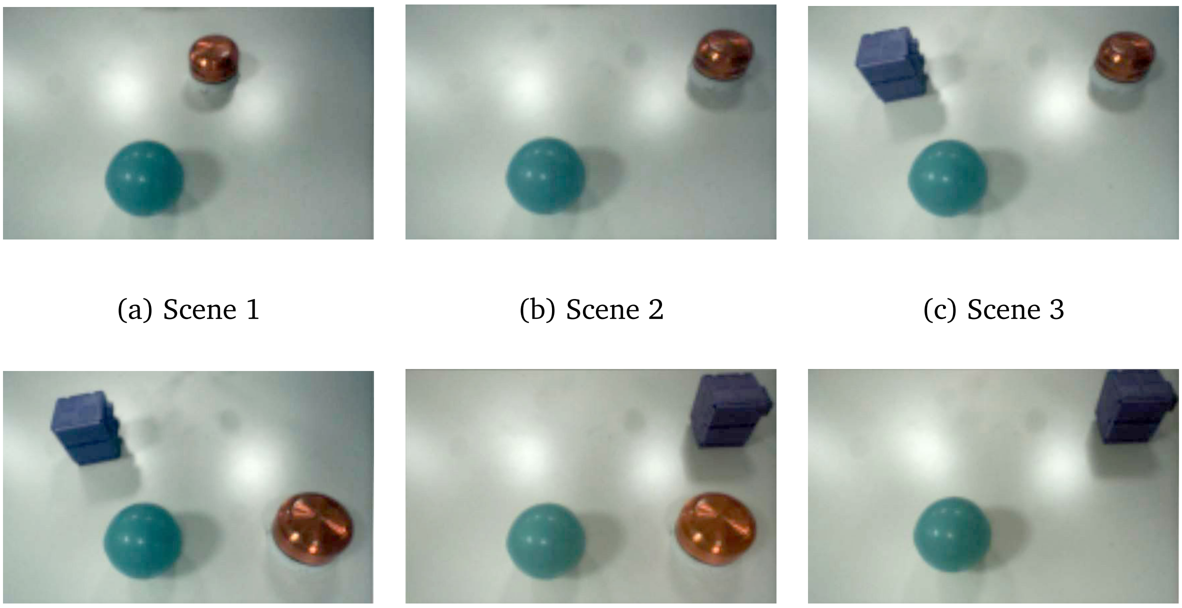
(e) Scene 5

(f) Scene 6

Figure 7.1: Experienced scene by Robot A

human-robot interaction module, where a human may inquire Baxter regarding details of experienced past events or even simply the familiarity status related to specific input, according to any of three cases introduced in Section 5.5.

Knowledge revision can be performed anytime during the first phase or even after the first phase is complete, which procedure is depicted in Figure 5.5. Known objects are associated with a label and multiple tags. The relationship among tags, SM item and episode is shown in Table 7.1. For example, during the first scene segmentation of Robot A as shown in Figure 7.1, two unidentified objects labels with their position with respect to the Field of View are consolidated as episode 1. As time progresses, detected novel objects are identified and consolidated when necessary during the scene capture process. In scene 2, a tennis ball is introduced to Baxter, and its corresponding SM item is consolidated as well as the corresponding episode. In scene 3, no novel objects are detected. However, since the position of an object has changed, a new episode is consolidated to mark the changes of environment. The episode contains reference to the detected objects from Baxter's current knowledge of SM item. So far, Baxter has three episodes (episode 1, 2, and 3) and three unlabeled SM items with color representation and texture representation, with no tags associated to the SM items. This process continues until all



(a) Scene 1

(b) Scene 2

(c) Scene 3

(d) Scene 4

(e) Scene 5

(f) Scene 6

Figure 7.2: Experienced scene by Robot B

the witnessed scenes are processed. At this point, a knowledge revision process is commenced, where a human assigns a label and tags that are associated with the objects. After revision, the object labeled as sirenlamp is assigned the tags orange, and round. Since the EM and SM are interconnected, the label of the sirenlamp in episode 1, 2, and 3 are automatically updated from a less meaningful, random alphanumeric placeholder label. In a similar spirit to the experimental design in the experimental 1, to make things more interesting, some common tags (e.g., round, yellow, toy) are assigned to the known objects. The tags yellow and ball are assigned to the object tennisball, and the tags wood, brown, and cube are assigned to the cube object. The same procedure applies to the rest of the scenes captured, as shown in Figure 7.1. Please note that other details within the episodes such as object count, timestamp, scene label are omitted for clarity.

Case 1. First, both Robot A and Robot B are presented with the set of objects that are exclusively and mutually available for both robots one at a time, i.e., the objects listed in Table 7.1. This means that both robots will consolidate these scenes, where only one object is detected, during the interaction with human. In short, ***the robots gain knowledge during the interaction***. This phenomenon occurs when the robot is presented with novel objects during human-robot interaction, which is the fundamental

Table 7.1: The connection between tags, SM items, and episodes

(a) Object-tag associations after revision for Robot A

	Objects detected by Robot A			
Label Found in Tags	Sirenlamp Scene 1-3 orange round	Tennisball Scene 2-6 yellow ball	Cube Scene 6,7 wood brown cube	Softtoy Scene 1-4 black yellow toy

(b) Object-tag associations after revision for Robot B

	Objects detected by Robot B		
Label Found in Tags	Orangelamp Scene 1-5 siren lamp	Purplecube Scene 3-6 purple toy cube	Tealball Scene 1-6 round ball toy

purpose of ERIS: **to enforce mutual understanding between human and robot**. To verify the robot capabilities to exhibit object familiarity, six objects are presented to each robot one at a time: every time, the robot provides the familiarity status related to the presented object. For the scene familiarity, we rearranged the objects on the table to resemble specific experienced scenes and also set a completely new arrangement to see whether the witnessed scene is familiar to both robots.

Case 2. We query each robot about the familiarity status of contextual information via a command-line interface. For the object familiarity, this will only correlate with known SM item. In this case, we expect a Boolean response for the familiarity status, with the addition of objects that are similar to the context. For the scene familiarity, the input is a context item as defined in Definition 11 and will correlate with either SM, EM, or even both, depending on the context. Similar to the previous experiment, to achieve a more variety of inputs, a number of simple mapping indicators were implemented, such as leftmost and rightmost.

Case 3. For the object familiarity, several objects are presented one at a time, with additional contextual information regarding the object characteristics that we would like each robot to remember. For the scene familiarity, the workspace is rearranged in various configurations and a scene is captured for each configuration as the input for each individual robot with additional contextual information.

Finally, for the more advance memory recollection procedure, a mixed of common and different questions will be posed to both robots. According to the working hypothesis of The Principle of Subjectivity, we expect both robots to respond in a different fashion.

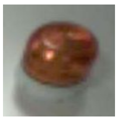
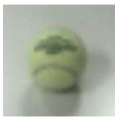
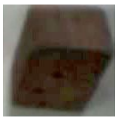
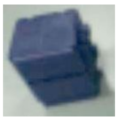
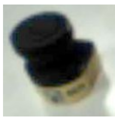
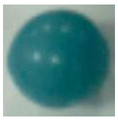
7.2 Experimental Evaluation

7.2.1 Results

The following subsection is a summary of the experimental results, and the next subsection is dedicated to the major discussion of the result. The results for case 1 are shown in Table 7.2. Six objects were presented to both Robot A and Robot B one at a time to test the object familiarity, and a human inquired the familiarity status of the object. For the scene familiarity, the workspace was rearranged as the depicted configurations, and a human inquired Baxter the familiarity of the scene, which eventually covers the familiarity of all objects within the scene. The results for case only involved Boolean responses. Table 7.3 shows the essential results for case 2. Aside from Boolean responses for the context in Table 7.3a, details about objects that support the Boolean response or similar to the desired input are provided by the query server module of the corresponding robot. Table 7.4 and Table 7.5 presents both object and scene familiarity results for case 3, respectively. For the object familiarity applies to case 1 and case 3, since the criterion for similarity matching is based on color and texture features, objects presented from different viewing angle sometimes may be recognized as novel objects, due to the insufficient recognition rate with respect to the predefined threshold, which in this case 70% for both color and texture recognition. For simplicity, those multiple SM items which refer to the same object will be treated as one SM item. To solve this problem, a robust incremental learning based visual tracking will be integrated as one of the immediate future works.

Table 7.2: Experimental results for Case 1

(a) Object familiarity test

Robot	Is object familiar?					
						
A	✓	✓	✓	✗	✓	✗
B	✓	✗	✗	✓	✗	✓

(b) Scene familiarity test

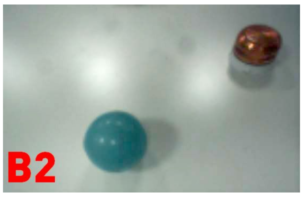
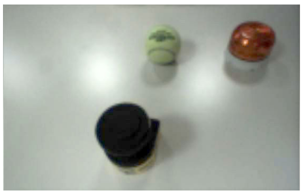
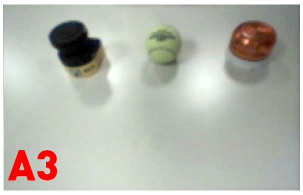
Robot	Is scene familiar?		
			
A	✗	✗	✓
B	✓	✗	✗

Table 7.6 shows the results of the advance memory recollection feature given the current knowledge possessed by each robot after experiencing all respective past events.

7.2.2 Discussion

Let us discuss the results for case 1. For the object familiarity test, four objects are able to be identified by Robot A (which is sirenlamp, tennisball, cube and softtoy), and three objects by Robot B (purplecube, tealball, and orangelamp). The unfamiliar objects are due to the absence of the corresponding SM item. For the scene familiarity, three different arrangements of the workspace were set up for both robots. Robot A and Robot B are able to recognize one scene each to be familiar, depending on their past experience.

Another thing to be clarify, during the recognition of objects within a scene, the perspective of the object affects the familiarity result. There are cases where, for instance, a cube that is positioned $45°$ to the left and $45°$ to the right might be recognized as two different objects. As a temporary workaround, as we mentioned before, here we treat them

Table 7.3: Experimental results for Case 2

(a) Object familiarity with suggestions

Context	Robot A		Robot B	
	Familiar	Suggestion	Familiar	Suggestion
Lamp	✓	Sirenlamp	✓	Orangelamp
Toy	✓	Softtoy	✓	Tealball, Purplecube
Yellow Toy	✓	Softtoy	✗	Tealball Purplecube
Purple object	✗	-	✓	Purplecube
Round entity	✓	Sirenlamp	✓	Tealball
Brown object	✓	Cube	✗	-

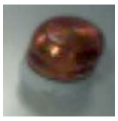
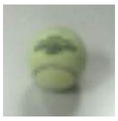
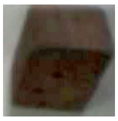
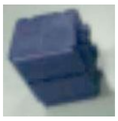
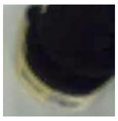
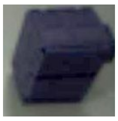
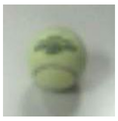
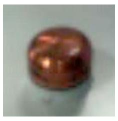
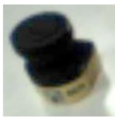
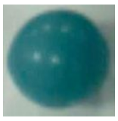
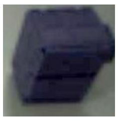
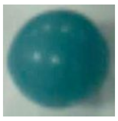
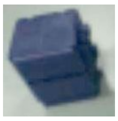
(b) Scene familiarity with corresponding scenes

Given Context	Familiar Scene	
	Robot A	Robot B
a ball at the leftmost	Scene 2,5-6	Scene 1-2,5-6
a ball with a cube	Scene 6	Scene 6
at least, a ball with a cube	Scene 6	Scene 3-6
yellow toy	Scene 1-4	✗
a lamp and 1 other object	Scene 1	Scene 1-2
a cube and 2 other objects	✗	Scene 3-5
2 objects with a cube at the rightmost	Scene 6	Scene 6

as the same object by assigning the same tags and label. This consequently may results in multiple SM items that refers to one particular object. So far, both robots are able to recognize familiar situations based on a single input, either an image view of a single object or witnessing a particular workspace arrangement. As two different robots with different experience, they already shown different response based on the experienced past events. We will see this more clearly in the subsequent results.

With a different approach, the input for case 2 is only constituted by a lexical context instead of a direct visual stimuli. For instance, Baxter is provided with verbal inputs instead of presenting an object. The lexical input *"are you familiar with a lamp?"* is provided to both robots. This affects the overall processing of memory recollection, as the color feature and texture feature will not be very much useful, and only tags related

Table 7.4: Experimental results for Case 3

Robot	Given a particular context, is the object familiar?					
	 + ball	 + orange	 + purple	 + yellow	 + toy	 + toy
A			✗	✗		✗
B		✗		✗		

to each SM item will determine the robots responses. We argue that **this phenomenon emphasizes the usefulness of tags**, as information such as color and texture can be encoded to memory but will not be much of use for the decoding other than familiarity matching. Tags associated to the SM item from human-robot interaction will help the robot infer the lexical query since there is no visual feed involved during the interaction. This can be analogically compared to the different mental representation possessed by human when dealing with different kinds of stimuli (e.g., visual or lexical stimuli), as elaborated by Kosslyn (2005) and based on the evidence by Tomasino, Werner, Weiss, & Fink (2007).

For the object familiarity of lexical stimuli, the robot tries to identify the object from both object label and associated tags. The context are listed in Table 7.3a based on the formal design in Definition 11. For the context of lamp, both robots successfully recognize the desired object as a familiar object, due to the fact that after interaction with human, the object label or the tag contains a lexical information of lamp. However, due to the previous interaction through the knowledge revision, each robot yields different response of the object label, regardless that they refer to the same actual object. For the same reason, the context toy also successfully recognized. However, both robots refer to different objects that they know to be related to the keyword toy. When the robots were asked about yellow toy, Robot A recognized softtoy as to be closely related with *yellow toy* as both tags yellow and toy are associated with the object softtoy. On the other hand, Robot B is not familiar with the context, but other objects associated with the the tags is suggested, i.e., tealball and purplecube. The rest of the given context related to the queried object applies the same procedure.

With the first context is given, Robot A is familiar with the desired scene as of scene 2, 5, and 6, and Robot B with the scene 1, 2, 5, and 6. Considering the definition of *event*, the architecture may interpret this result one step further by inferring that, for example, an *event A* and *event B* are events where a ball is located at the leftmost of the field of view, which occurrence is between scene 5 and 6 for the event A experienced by Robot A, and between scene 1 and 2 for the event A experienced by Robot B. Similarly *event B* experienced by Robot B occurs between scene 5 and 6. Now the remaining scene for Robot A is scene 2, which does not have a subsequent scene to be defined as an event. If we briefly recall the definition of episode and event as defined in Section 9, an event consists of multiple, timestamp-ordered episodes during the period of that event, and an episode is a digest of a scene. In order to define an event, multiple episodes with respect to the desired criterion must be present. Therefore, the event of a ball located at the leftmost cannot be defined with the presence of episode 2 solely. As previously discussed, an event is *dynamic*, meaning that changes within the Field of View are to be expected during the occurrence of an event. On the contrary, a scene (and its corresponding episode) is *static*. Therefore, we can consider that at a certain point of time (i.e., at scene 2), a workspace configuration matched the criterion.

The second question of case 2 scene familiarity is about the familiarity of a scene where *a ball and a cube are identified*. Both Robots yields only scene 6, which means only a matching workspace configuration exists instead of an event, considering that scene 6 for both robots are totally different. The third given context for the next question is about a scene where *at least a ball and a cube is identified*. Robot A recognizes scene 6 as the matched scene for the criterion, and Robot B recognizes scene 3 until scene 6.

If we recall the *event B* for Robot B in the first question of case 2 that event B consists of scene 5 and 6, now we have scene 3 until scene 6 defined as a different event. As two different events, scene 5 and scene 6 overlap for the two given context. This simultaneously highlights the **flexible** and **subjective** property of an event. Furthermore, since the definition of events depends on the experienced visual stimuli, both Robot A and Robot B may possess a completely different definition of *event A*. This emphasized the **personal** property of an event. The rest of the given context for case 2 applies the same procedure. From these results so far, Baxter shows the capability of recalling its personal past events given only contextual information, and the results clarify the properties of events: flexible, subjective and personal.

Case 3 involves both visual stimuli and contextual information as the input. For the object familiarity, presented objects and the given context are shown in Table 7.4. Six presented objects are paired with various contexts to see how familiarity is affected by the context for both robots. For instance, the orangelamp is presented with an additional context of ball and yields the tennisball as the object that is also associated with ball category for Robot A and tealball which associated with round category for Robot B. This is due to the difference of the interaction history for both robots. Since the result of the current interaction depends on the past interaction, current robot knowledge and contextual information, it is possible for both robots to yields the current interaction result in a various, interesting manner. Other examples of unfamiliar results yielded (such as cube, purplecube for Robot A, and tennisball, purplecube for Robot B) when the presented objects do not recognized as a familiar object with the given context.

Table 7.5a and Table 7.5b show two distinct workspace arrangements monitored by each robot, and additional contexts are given to check the familiarity status. Unfamiliar workspace arrangement with a provided context were also tested to see both robots' capability to remember the desired past snapshots based on the present state of the stimuli. It turns out that given an unfamiliar workspace (i.e., the middle scene in Table 7.5a) and supplemented with the right context, Robot A is capable of recognizing the desired scene based on a stimuli that has never been experienced before. On the contrary, given a familiar scene and an unsuitable context may yields unfamiliar result, as demonstrated by Robot B with the middle scene in Table 7.5b. These result demonstrates the influence of context through the personal scene recollection process.

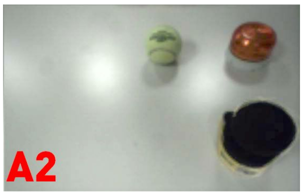
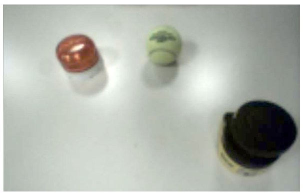
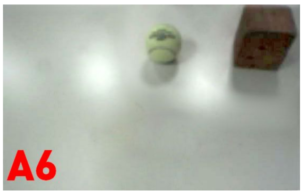
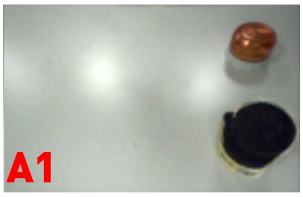
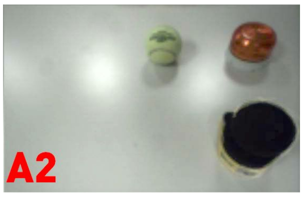
Next, from Table 7.6, we discuss the more advance memory recollection process that allows a more dynamic interaction as one of the features of ERIS. Both Robot A and Robot B have been queried with questions that not only yields Boolean results, but also capable of retrieving information from the contents of experienced scenes/events and general facts from the SM data in the LTM. Both robots were asked a simple question (e.g., *"What is this object?"* by presenting the object to each robot), to a more difficult questions (e.g., *"Did you move the object?", "How many objects have you seen so far?"*).

Both robots were given several identical questions, and different questions regarding their past experience. For instance, by presenting the object lamp to the two robots and posing a question, both robots are able to respond and give a more derived results regarding the question, instead of just returning Boolean response as in Table 7.2a and Table 7.4. The results for both robots are pretty interesting, since the current robot knowledge depends on the interaction history, both robots yielded different results although queried several identical questions. This makes robots integrated with ERIS achieve a dynamic and more personal nature of epigenetic robots in terms of interaction quality.

This concludes our analysis about the manifestation of Baxter's past experience and possessed knowledge, distinguished as two different robots. We showed how different interaction history and contextual information affects the current Human-Robot interaction, which corroborates the principle of subjectivity pointed out by Stoytchev (2009).

Table 7.5: Scene familiarity

(a) Scene familiarity test for Robot A

Robot	Is scene familiar?		
	 + no ball	 + sirenlamp on the right side	 + with softtoy on the left side
A			✗

(b) Scene familiarity test for Robot B

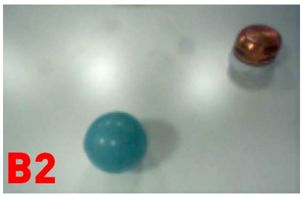
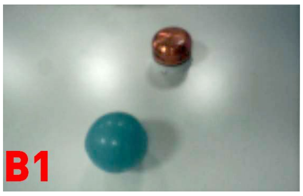
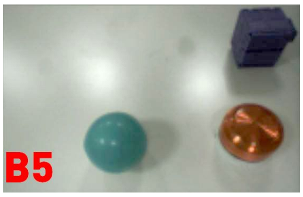
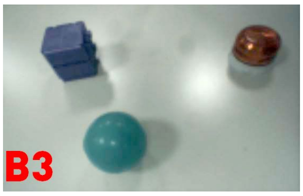
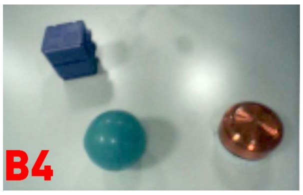
Robot	Is scene familiar?		
	 + with purple cube	 + with purplecube, no orangelamp	 + with purplecube on the left side
B		✗	

Table 7.6: Advance memory recollection

(a) Questions and visual stimuli exposed to Robot A

Question for Robot A	Answer
What is this?	Sirenlamp
What cube were you presented with?	Cube
How many toy objects have you seen?	1
How many yellow objects have you seen so far?	2
Did you remove any of the yellow objects?	Yes, tennisball softtoy
How many objects left after you remove the tennis ball?	1, cube
How many objects, at least, were in the workspace?	1, Scene 5

(b) Questions and visual stimuli exposed to Robot B

Question for Robot B	Answer
What is this?	Orangelamp
What cube were you presented with?	purplecube
How many toy objects have you seen?	2, tealball, purplecube
Did you move the tealball?	No
How many yellow objects have you seen so far?	0
What objects did you move?	orangelamp purplecube

Chapter 8

Summary, Contributions, and Future Work

> I not only use all the brains that I have, but all that I can borrow.
>
> Woodrow Wilson

In this chapter, we provide summary of this dissertation and the contribution to both of the epigenetic robotics, and to robotics community in general. We then present several interesting directions for future work, and some final words.

8.1 Dissertation Summary

Developmental/epigenetic robotics is a relatively new paradigm as well as an interdisciplinary branch of research area that revolves around robotics, cognitive psychology, and behavioral science; where the research area investigates viable methods to make a robot self-develop its knowledge, or even personality. This dissertation addressed the problem of mitigating the interaction gap present during Human-Robot Interaction (HRI), considering the robot is not designed under developmental/epigenetic paradigm. To do so, we develop an epigenetic architecture called Epigenetic Robot Intelligent System (ERIS).

The development of ERIS is motivated by the lack of available open-source epigenetic platform that achieves the features of ERIS. Similar epigenetic architectures available and cognitive architectures are explicitly distinguished and discuss thoroughly in Chapter 2, as well as the significance of contextual information within HRI. We then introduce the fundamentals of human memory in Chapter 3, which provides insight about the four phases of memory processing (memory formation, consolidation, recollection, and revision) in human brain based on the state-of-the-art studies in developmental psychology.

We elaborated the implementation details of ERIS as a ROS stack in Chapter 4, discussed each ROS package and their relationship to the details presented back in Chapter 3. We presented the conceptual explanations of familiarity mechanism within the implemented ERIS stack in Chapter 5.

Two separate experiments were conducted to validate the following topics: (1) manifestation of robot personal experience; and (2) the Principle of Subjectivity. Both experiments are elaborated in Chapter 6 and Chapter 7, respectively. The first experiment validates the capability of self-developed knowledge exhibited by the robot by interacting with human or the environment. The validation is further expanded in the second experiment, from the perspective of an interesting principle called *The Principle of Subjectivity*, which fundamentally highlights the influence of human interaction history and received stimuli to robot personal experience and knowledge development. At each experiment, we evaluate the results and provide related discussions.

8.2 Contributions

This thesis contributes the development of ERIS, an epigenetic robot framework. As the contributions to the epigenetic/developmental robotics field, ERIS allows robots to exhibit the following two major phenomena:

- progressively self-develop their knowledge involving their unique, past experience
- “remember”/recall relevant past events happening in the environment during robot-environment and human-robot interaction.

Based on the state-of-the-art studies from developmental psychology and behavioral science field, the following are additional contributions to the field:

- Based on the notion of **memory item** as the building block, and with the emphasis on the interconnectivity between memory components presented through an explicit formalism of ERIS contributes to the capability of consolidating and recalling past experiences in the form of events and general knowledge (which represent Episodic Memory and Semantic Memory).
- the ability that allows robots to deal with verbalized contextual information as the result of interaction with the external world, including humans, to achieve the most

natural way of interaction possible. The incorporation of contextual information is systematically elaborated in the formal design of ERIS.

- ERIS complies with *dual-coding theory* in stimuli processing, where verbal and imagery processing have different protocol.

- ERIS demonstrated to be complies with *the Principle of Subjectivity* when integrated in two identical robots, yields in personalized robot experience and knowledge through interaction.

- Robot Operating System (ROS) implementation of ERIS to ensure high compatibility of integration in ROS-compliant general purpose robots. Also the fact that ROS is an open source platform, ERIS is also a contribution to the robotics society, which is highly accessible and extensible by the community. The source code of ERIS is available at `https://github.com/ferdianap/eris`.

8.3 Future Work

Despite the given contribution and the progress made to the developmental robotics field, the work in this dissertation is a small step towards the realization of a fully epigenetic robotic system and more personalized robot companion, including the exhibition of dynamic robot personality and individuality. Several interesting research directions are available to be explored, some of them are the following:

- **Modeling the procedural memory**

  This enhancement allows the robot to recall its motor skills performed at a particular time/event in the past. It is a challenging task to have a single universal procedural memory design that suits a variety of robots with different physical structure.

- **The user's and robot's personality identification**

  This area requires an explicit formalism of *personality* to ensure that proper estimation can be done. At our current progress, the notion of robot personality originated from the manifestation of robot personal experienced is rather vague. One viable idea is to develop a piece of software that manages the classification of human behavior and personality traits out of the well-known profiling method by Korem (1997). This profiling technique allows us to profile people on the spot, or even profile people before meeting the individual. Despite a powerful technique, this will requires us to develop several things before reaching this stage, such as:

- a human gestures recognition module for ERIS,
- a sophisticated natural language processing module to analyze human verbal communication.

- **The robot's personality adaptation to changes to human's mood and context**

  A direct consequence of having a robot understanding an interacting human's personality is to be able to adapt to changing *moods*, assuming that we consider mood as a significant part of one's personality. This implies to be able to estimating human's personality online and to provide appropriate responses. Once this is available, we may think of estimating not only profiles in a static way (e.g., the human has a "sergeant-manager" profile has a tendency to perform a list of actions), but to identify trends in viable actions corresponds to (or even contradicts) the profile (e.g., "typically, the human I am interacting with is angry in the morning, but not this morning due to several reasons"), in order to provide a sort of proactive assistance, and more importantly, how to appropriately deal with all the possible actions that might be performed by the human with such profiles.

- **How robot's personality affect the way it plans and behaves**

  This is another apparent consequence in the robot having its own personality. Personality affects how we interpret the environment, plan and act in the real world. As such, we may identify models to tune perception and action-oriented robot behaviors, for instance, in the case of adapting planning algorithms. Typically, robot plans are achieved (using, for instance, STRIPS-like algorithms) according to some sort of optimal criteria (i.e., reducing the number of steps). However, this is not how humans plan, which is sub-optimal and depends on experience (e.g., I may re-adapt previously successful plans, or discard an optimal plan because it is way similar to a plan that failed in the past). Interestingly, this is closely related with the directive property of autobiographical memory, discussed in Section 3.4, and since memory affects personality (at least in human), the way a robot modulates planning may be connected to its personality.

- **Interaction of robots having different personalities**

This is a bit different from the usual stuff, but it is very interesting nonetheless, in the sense that it possibly grounds experiments in psychology. The goal is to understand how two agents with different personalities interact in doing a common task, obviously involving interaction or competition of some sorts (e.g., assembling something together). Since we can evolve robot personalities by providing robots with proper stimuli, we can precisely define how personalities affect the interaction, also in simulation if necessary. This may eventually ground work on so-called *dyads*, relationships between couples of agents in interaction, also with insights in interactions with humans.

- **Exploitation of "knowledge transfer"**

  A knowledge gain by a robot might be reusable at a certain degree by another robot that has different physical structure. The reusable aspects includes an adaptation of motor movement planning, for instance, a scenario where a recipient robot with a set of transferred knowledge recalls a memory which relates to an unfeasible motor skills to the recipient robot, but feasible to the donor robot. Although it seems that only the robot-independent memories will be influenced (i.e., Semantic Memory and Episodic Memory), the feasibility of the Procedural Memories between the recipient-donor robot can be used as an advantage to provide an alternative possible motor movements using the set of existing motor skills possessed by the recipient robot.

- **Designing a metric for measuring/estimating both human and robot personality**

  This feature benefits from both robot and human perspective. Human may have a better interaction experience when the robot has a distinct personality, and robot can estimate human personality (to "profile" them), determine its own personality, and change it accordingly. To achieve this, some significant improvements that needs to be done including:

  – integrating ERIS stack with a speech-based dialog system grounded with respect to the cue-value pair based formalism;

  – integrating a state-of-the-art natural language processing module;

  – improving the ViSor package with incremental-learning vision tracking module, i.e., the work by Ross, Lim, Lin, & Yang (2008); and

  - integrating modules specifically for analyzing psychological phenomenon of interaction, e.g., recognizing human posture, gesture/body language, and emotion.

- **Knowledge generalization**

  Currently, our system represents knowledge in a "flat" form, using many $\langle$cue, value$\rangle$ pairs. This representation is basic, has obvious benefits but also drawbacks. The first drawback is that it doesn't capture the natural top-down and bottom-up nature of knowledge, which continuously abstract grounded knowledge and grounds abstract knowledge, as reflected by the spreading activation model by Collins & Loftus (1975) and hierarchical structure of semantic memory by Collins & Quillian (1969). It is also a good idea to combine the benefits of both models, and we may further consider grounding the representation of the environment (as well as other traits of robot's personality) at different levels.

There are other aspects of epigenetic robotics that needs to be explored, such as mentioned by Stoytchev (2009):

- **The Verification Principle**

  A robot should be able to create and maintain knowledge that can be verified by itself. As the first researcher in AI who explicitly mention about this profound principle, Sutton (2001) once claimed that *"the key to a successful AI is that it can tell for itself whether or not it is working correctly"*. The term *verification* here refers to automatic, relative validation of the knowledge by the robot and for the robot. In principle, he stated that anything that is potentially learnable must be verifiable in an autonomous manner. This is what is argued by Stoytchev (2009) that researchers (or in particular, programmers) must carefully design the structures for the quantification of the abstract knowledge, so that the potentially learnable material (in an abstract form) can be quantified and verified.

- **The Principle of Embodiment**

  A robot should be able to dynamically distinguish its own body parts from the environment. Although this topic is quite popular in the field of developmental psychology (Rochat, 2003; Ramachandran, Blakeslee, & Sacks, 1998; Iriki, Tanaka, & Iwamura, 1996) and there have been a progress regarding this topic (Stoytchev, 2003), there are still more aspects to be considered and explored from the epigenetic robotics perspective.

- **The Principle of Grounding**

  Similar to the Symbol Grounding problem by Harnad (1990), this principle is considered from epigenetic robotics perspective by Stoytchev (2009), where temporal contingency is considered as an approach to deal with this challenge. However, this is only the beginning and there could be a lot more to be explored for this topic.

- **The Principle of Incremental Development**

  This principle is considered to be the generalized domain of incremental learning as a specific machine learning technique that solves a specific problem. Although this principle is influenced by the other principles, only a small portions of conceptual progress achieved so far. Having a formalism and implementation based on this principle is considered a milestone in the research area.

8.4 Conclusions

This dissertation work addresses the problem of epigenetic robotics, where a robot should be able to developmentally "grows" in terms of robot knowledge, and the manifestation of robot personal experience. We have achieved this by designing an epigenetic architecture to accommodate information gathered from the interaction with the world, including humans.

Chapter 3 provides the details regarding high-level memory and Chapter 4 elaborates the architecture implementation as well as the interconnectivity between each component (including memory, which concept is discussed in Chapter 3). We have demonstrated that using ERIS, a ROS compliant general purpose robot, i.e., Baxter Research Robot used in the experiments are able to manifest personal experience in a very fundamental forms. In our current progress, only visual stimuli are considered. We took the experiment further by exploiting *The Principle of Subjectivity*, as one of the most significant principles of epigenetic robotics, by analyzing Baxter as two different robots in two independent, separate experiments. In each experiment, Baxter is exposed to different visual stimuli and different interaction history. The two aspects influence the robot personal development, validated through verbal interactions with human, where query-based question regarding the past experience may be posed to Baxter.

Overall, this research work suggests that when a robot has a formal and explicit model of knowledge which design to adapt to human knowledge, we can expect robot to progressively self-develop its knowledge through interaction. Furthermore, not only contextual information influence human interaction with others, it also enhance the interaction quality between robot and human.

References

Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research*, *812*(1–2), 23–37.

Allen, C., & Bekoff, M. (1999). *Species of mind: The philosophy and biology of cognitive ethology*. Cambridge, MA, USA: MIT Press.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832–843.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive science*, *26*(1), 39–83.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive psychology*, *30*(3), 221–256.

Antonelli, M., Gibaldi, A., Beuth, F., Duran, A., Canessa, A., Chessa, M., . . . Sabatini, S. (2014, Dec). A hierarchical system for a distributed representation of the peripersonal space of a humanoid robot. *IEEE Transactions on Autonomous Mental Development*, *6*(4), 259–273.

Atkinson, R., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Baddeley, A. (1987). But what the hell is it for? In M. Grueneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 3–18). UK: John Wiley.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Section A*, *49*(1), 5–28.

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, *4*(05), 523–526.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.*, *4*(11), 417–423.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1–29.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2014). *Memory* (2nd ed.). Psychology Press.

Baddeley, A., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 8, pp. 47–89). USA: Academic Press.

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737–1743.

Baddeley, A. D. (1966a). The influence of acoustic and semantic similarity on long-term memory for word sequences. *The Quarterly journal of experimental psychology*, *18*(4), 302–309.

Baddeley, A. D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, *18*(4), 362–365.

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*(1), 126.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575–589.

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 193–243). Cambridge, UK: Cambridge University Press.

Bartlett, F. C. (1932). Remembering: An experimental and social study. *Cambridge: Cambridge University*.

Bastan, M., Cam, H., Gudukbay, U., & Ulusoy, O. (2010). Bilvideo-7: An mpeg-7- compatible video indexing and retrieval system. *IEEE Multimedia*, *17*(3), 62–73. doi: http://doi.ieeecomputersociety.org/10.1109/MMUL.2009.74

Bellas, F., & Duro, R. (2004). Multilevel darwinist brain in robots: Initial implementation. In *Proc. int. conf. inf. cont. autom. robot (icinco)* (pp. 25–32).

Bellas, F., Faina, A., Varela, G., & Duro, R. J. (2010, July). A cognitive developmental robotics architecture for lifelong learning by evolution in real robots. In *Proc. int. joint conf. neural networks* (pp. 1–8). Barcelona, Spain.

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in cognitive sciences*, *15*(11), 527–536.

Blank, D., Kumar, D., Meeden, L., & Marshall, J. B. (2005). Bringing up robot: Fundamental mechanisms for creating a self-motivated, self-organizing architecture. *Cybernetics and Systems: An International Journal*, *36*(2), 125–150.

Bliss, J. C., Crane, H. D., Mansfield, P. K., & Townsend, J. T. (1966). Information available in brief tactile presentations. *Perception & Psychophysics*, *1*(4), 273–283.

Bluck, S. (2003). Autobiographical memory: Exploring its functions in everyday life. *Memory*, *11*(2), 113–123.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive psychology*, *11*(2), 177–220.

Brewer, W. F. (1986). What is autobiographical memory? *Autobiographical memory*.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 207–230.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and autonomous systems*, *6*(1), 3–15.

Bruusgaard, J., Liestøl, K., Ekmark, M., Kollstad, K., & Gundersen, K. (2003). Number and spatial distribution of nuclei in the muscle fibres of normal mice studied in vivo. *The Journal of Physiology*, *551*(2), 467–478.

Bruusgaard, J. C., Johansen, I. B., Egner, I. M., Rana, Z. A., & Gundersen, K. (2010). Myonuclei acquired by overload exercise precede hypertrophy and are not lost on detraining. *Proceedings of the National Academy of Sciences*, *107*(34), 15111-15116. Retrieved from `http://www.pnas.org/content/107/34/15111.abstract` doi: 10.1073/pnas.0913935107

Budiu, R., & Anderson, J. R. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science*, *28*(1), 1–44.

Bullemer, P., Nissen, M. J., & Willingham, D. B. (1989). On the development of procedural knowledge. *Journal of Experiemental Psychology: Learning, Memory and Cognition*, *15*(6), 1047–1060.

Byrne, M. D., & Kirlik, A. (2005). Using computational cognitive modeling to diagnose possible sources of aviation error. *The international journal of aviation psychology*, *15*(2), 135–155.

Cangelosi, A., & Schlesinger, M. (2014). *Developmental robotics: From babies to robots*. The MIT Press.

Carlson, N. R., Heth, D., Miller, H., Donahoe, J., & Martin, G. N. (2009). *Psychology: the science of behavior*. Pearson.

Chaigneau, S. E., Barsalou, L. W., & Zamani, M. (2009). Situational information contributes to object categorization and inference. *Acta Psychologica*, *130*(1), 81–94.

Chastagnol, C., Clavel, C., Courgeon, M., & Devillers, L. (2014). Designing an emotion detection system for a socially intelligent human-robot interaction. In *Natural interaction with robots, knowbots and smartphones* (pp. 199–211). Springer.

Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews*, *26*(2), 105–125.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological review*, *82*(6), 407.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, *8*(2), 240–247.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British journal of psychology*, *55*(4), 429–432.

Conway, M. A. (2005). Memory and the self. *Journal of memory and language*, *53*(4), 594–628.

Conway, M. A., & Bekerian, D. A. (1987). Organization in autobiographical memory. *Memory & Cognition*, *15*(2), 119–132.

Conway, M. A., Cohen, G., & Stanhope, N. (1992). Very long-term memory for knowledge acquired at school and university. *Applied cognitive psychology*, *6*(6), 467–482.

Conway, M. A., & Holmes, E. (2005). Autobiographical memory and the working self. In N. Braisby & A. Gellatly (Eds.), *Cognitive psychology* (pp. 507–43).

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261–288.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, *19*(4), 450–466.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*(2), 255–267.

Dayoub, F., Duckett, T., Cielniak, G., et al. (2010). Toward an object-based semantic memory for long-term operation of mobile service robots. *Workshop Semantic Map. Auton. Knowledge Acquis. (IROS)*.

Denei, S., Mastrogiovanni, F., & Cannata, G. (2015). Towards the creation of tactile maps for robots and their use in robot contact motion control. *Robotics and Autonomous Systems*, *63*, 293–308.

Dodd, W. (2005). *The design of procedural, semantic, and episodic memory systems for a cognitive robot* (Unpublished doctoral dissertation). Vanderbilt University.

Dodd, W., & Gutierrez, R. (2005). The role of episodic memory and emotion in a cognitive robot. In *Proc. ieee int. workshop robot hum. commun. (roman)* (pp. 692–697).

Dritschel, B. H., Williams, J., Baddeley, A. D., & Nimmo-Smith, I. (1992). Autobiographical fluency: A method for the study of personal memory. *Memory & Cognition*, *20*(2), 133–140.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. Oxford University Press.

Emerson, E. A., & Halpern, J. Y. (1986). “sometimes” and “not never” revisited: on branching versus linear time temporal logic. *Journal of the Association of Computing Machinery*, *33*(1), 151–178.

Fernaeus, Y., H**r**akansson, M., Jacobsson, M., & Ljungblad, S. (2010). How do you play with a robotic toy animal?: a long-term study of pleo. In *Proceedings of the 9th international conference on interaction design and children* (pp. 39–48).

Fivush, R., Haden, C., & Reese, E. (1996). Remembering, recounting, and reminiscing: The development of autobiographical memory in social context. In D. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 341–359). UK: Cambridge University Press.

Gibbs Jr, R. W. (2005). *Embodiment and cognitive science*. Cambridge University Press.

Gilson, E. Q., & Baddeley, A. (1969). Tactile short-term memory. *The Quarterly journal of experimental psychology*, *21*(2), 180–184.

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and brain sciences*, *20*(01), 41–50.

Glucksberg, S., & Jr., G. N. C. (1970). Memory for nonattended auditory material. *Cognitive Psychology*, *1*(2), 149–156.

Gobet, F., & Lane, P. C. (2010). The chrest architecture of cognition: The role of perception in general intelligence. In *Procs 3rd conf on artificial general intelligence.*

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., ... others (2005). Designing robots for long-term social interaction. In *Intelligent robots and systems, 2005.(iros 2005). 2005 ieee/rsj international conference on* (pp. 1338–1343).

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *Brit. J. Psychol.*, *66*(3), 325–331.

Gordon, A. M., Westling, G., Cole, K. J., & Johansson, R. S. (1993). Memory representations underlying motor commands used during manipulation of common and novel objects. *Journal of Neurophysiology*, *69*(6), 1789–1796.

Graesser, A. C. (1981). *Prose comprehension beyond the word*. Springer Science & Business Media.

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. (1979). Recognition memory for typical and atypical actions in scripted activities: Tests of a script pointer + tag hypothesis. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 319–332.

Graesser, A. C., & Nakamura, G. V. (1982). The impact of a schema on comprehension and memory. *The psychology of learning and motivation*, *16*, 59–109.

Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 503.

Groves, P. M., & Thompson, R. F. (1970). Habituation: a dual-process theory. *Psychological review*, *77*(5), 419.

Gundersen, K. (2011). Excitation-transcription coupling in skeletal muscle: the molecular pathways of exercise. *Biological Reviews*, *86*(3), 564–600.

Hadziselimovic, N., Vukojevic, V., Peter, F., Milnik, A., Fastenrath, M., Fenyves, B., ... Stetak, A. (2014). Forgetting is regulated via musashi-mediated translational control of the arp2/3 complex. *Cell*, *156*(6), 1153–1166.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1), 335–346.

Hart, D., & Goertzel, B. (2008). Opencog: A software framework for integrative artificial general intelligence. *Frontiers in Artificial Intelligence and Applications*, *171*, 468.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, *7*(14), 2325–2330.

Irish, M., Hornberger, M., Lah, S., Miller, L., Pengas, G., Nestor, P., ... Piguet, O. (2011). Profiles of recent autobiographical memory retrieval in semantic dementia, behavioural-variant frontotemporal dementia, and alzheimer's disease. *Neuropsychologia*, *49*(9), 2694–2702.

Jeong, S., Arie, H., Lee, M., & Tani, J. (2011). Neuro-Robotics study on integrative learning of proactive visual attention and motor behaviors. *Cognitive Neurodynamics*, *6*(1), 43–59.

Jockel, S., Weser, M., Westhoff, D., & Zhang, J. (2008). Towards an episodic memory for cognitive robots. In *Proc. of 6th cognitive robotics workshop at 18th european conf. on artificial intelligence (ecai)* (pp. 68–74).

Jockel, S., Westhoff, D., & Zhang, J. (2007, Dec). Epirome-a novel framework to investigate high-level episodic robot memory. In *Proc. ieee int. conf. robot. biomim. (robio)* (pp. 1075–1080).

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 656–674.

Kallaluri, N., Even, J., Morales, Y., Ishi, C., & Hagita, N. (2013, May). Probabilistic approach for building auditory maps with a mobile microphone array. In *Proceedings of the 2013 ieee international conference on robotics and automation (icra 2013)*. Karlsruhe, Germany.

Kan, I. P., Alexander, M. P., & Verfaellie, M. (2009). Contribution of prior semantic knowledge to new episodic learning in amnesia. *Journal of Cognitive Neuroscience*, *21*(5), 938–944.

Kasap, Z., & Magnenat-Thalmann, N. (2010). Towards episodic memory-based long-term affective interaction with a human-like robot. In *Proc. ieee int. symp. robot hum. interact. commun. (ro-man)* (pp. 452–457).

Kaster, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*(1), 315–341.

Kintsch, W., & Van Dijk, T. A. (1977). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 63–94.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, *21*(1), 1–6.

Kokinov, B. N. (1994). The dual cognitive architecture: A hybrid multi-agent approach. In *Ecai* (pp. 203–207).

Korem, D. (1997). *The art of profiling: Reading people right the first time*. International Focus Press.

Kosslyn, S. M. (2005). Mental images and the brain. *Cognitive Neuropsychology*, *22*(3-4), 333–347.

Koutstaal, W., Wagner, A., Rotte, M., Maril, A., Buckner, R., & Schacter, D. (2001). Perceptual specificity in visual object priming: functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*, *39*(2), 184–199.

Kuppuswamy, N. S., Cho, S. H., & Kim, J. H. (2006, Oct). A cognitive control architecture for an artificial creature using episodic memory. In *Proc. int. joint conf. sice-icase* (pp. 3104–3110).

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial intelligence*, *33*(1), 1–64.

Lebiere, C., & Anderson, J. R. (2008). A connectionist implementation of the act-r production system.

Lederman, S. J., & Klatzky, R. L. (2009). Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, *71*(7), 1439–1459.

Lee, K. M., Peng, W., Jin, S.-A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication*, *56*(4), 754–772.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, *5*(2), 291–308.

Leite, I., Martinho, C., Pereira, A., & Paiva, A. (2008). icat: an affective game buddy based on anticipatory mechanisms. In *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems-volume 3* (pp. 1229–1232).

Lewis, M. W., Milson, R., & Anderson, J. R. (1987). The teacher's apprentice: Designing an intelligent authoring system for high school mathematics. In *Artificial intelligence and instruction: Applications and methods* (pp. 269–301).

Loftus, E. F., & Suppes, P. (1972). Structural variables that determine the speed of retrieving words from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 770–777.

Logie, R. H. (1995). *Visuo-spatial working memory*. Psychology Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Manis, J. G., & Meltzer, B. N. (1978). *Symbolic interaction: A reader in social psychology*. Allyn and Bacon, Boston.

Manjunath, B. S., Salembier, P., & Sikora, T. (2002). *Introduction to mpeg-7: multimedia content description interface* (Vol. 1). John Wiley & Sons.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.*, *58*, 25–45.

Mastrogiovanni, F., Scalmato, A., Sgorbissa, A., & Zaccaria, R. (2011). Problem awareness for skilled humanoid robots. *International Journal of Machine Consciousness*, *3*(1), 91–114.

Mastrogiovanni, F., & Sgorbissa, A. (2012). A biologically plausible, neural-inspired planning approach which does not solve 'the gourd, the monkey, and the rice' puzzle. *Biologically Inspired Cognitive Architectures*, *2*, 77–87.

Mastrogiovanni, F., & Sgorbissa, A. (2013). A behavior sequencing and composition architecture based on ontologies for entertainment humanoid robots. *Robotics and Autonomous Systems*, *61*(2), 170–183.

Mehl, M. R., & Conner, T. S. (2012). *Handbook for research methods for studying daily life*. New York, MY, USA: The Guildford Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Milner, B. (1962). Les troubles de la memoire accompagnant des lesions hippocampiques bilaterales. *Physiologie de l'hippocampe*.

Miwa, H., Umetsu, T., Takanishi, A., & Takanobu, H. (2001). Robot personality based on the equations of emotion defined in the 3d mental space. In *Robotics and automation, 2001. proceedings 2001 icra. ieee international conference on* (Vol. 3, pp. 2602–2607).

Morse, A. F., Belpaeme, T., Cangelosi, A., & Smith, L. B. (2010). Thinking with your body: Modelling spatial biases in categorization using a real humanoid robot. In *Proc. of 2010 annual meeting of the cognitive science society. portland, usa* (pp. 1362–1368).

Morse, A. F., de Greeff, J., Belpeame, T., & Cangelosi, A. (2010). Epigenetic Robotics Architecture (ERA). *IEEE Trans. Auton. Mental Develop.*, *2*(4), 325–339.

Murdock, B. (1967). Auditory and visual stores in short term memory. *Acta Psychologica*, *27*, 316–324.

Nakamura, G. V., Graesser, A. C., Zimmerman, J. A., & Riha, J. (1985). Script processing in a natural situation. *Memory & Cognition*, *13*(2), 140–144.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Neisser, U. (1986). Nested structure in autobiographical memory.

Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical psychology*, *1*(1), 35–59.

Nelson, K. (2003). Self and social functions: Individual autobiographical memory and collective narrative. *Memory*, *11*(2), 125–136.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: a social cultural developmental theory. *Psychology Review*, *111*(2), 486–511.

Nigro, G., & Neisser, U. (1983). Point of view in personal memories. *Cognitive Psychology*, *15*(4), 467–482.

Norman, D. A., & Shallice, T. (1986). *Attention to action*. Springer.

Nuxoll, A. M. (2007). *Enhancing intelligent agents with episodic memory* (Unpublished doctoral dissertation). University of Michigan.

Nuxoll, A. M., & Laird, J. E. (2004, July). A cognitive model of episodic memory integrated with a general cognitive architecture. In *Proceedings of the 2004 ieee international conference on cognitive modeling (iccm 2014).* Pittsburgh, Pennsylvania, USA.

Nuxoll, A. M., & Laird, J. E. (2012). Enhancing intelligent agents with episodic memory. *Cognitive Systems Research*, *17*, 34–48.

Parmiggiani, A., Maggiali, M., Natale, L., Nori, F., Schmitz, A., Tsagarakis, N., ... Metta, G. (2012). The design of the icub humanoid robot. *International Journal of Humanoid Robotics*, *9*(04).

Peters II, R. A., Hambuchen, K. A., & Bodenheimer, R. E. (2009). The sensory ego-sphere: a mediating interface between sensors and cognition. *Autonomous Robots*, *26*, 1–19.

Phillips, J. L., & Noelle, D. C. (2005). A biologically inspired working memory framework for robots. In *Proc. ieee int. workshop robot hum. commun. (roman)* (pp. 599–604).

Pillemer, D. (2003). Directive functions of autobiographical memory: The guiding power of the specific episode. *Memory*, *11*(2), 193–202.

Pillemer, D. B. (1998). *Momentous events, vivid memories*. Harvard University Press.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., & Petersen, S. E. (2012). The attention system of the human brain: twenty years after. *Annual Review of Neuroscience*, *35*, 73–89.

Prince, C., Helder, N., & Hollich, G. (2005). Ongoing emergence: A core concept in epigenetic robotics.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., ... Ng, A. Y. (2009). Ros: an open-source robot operating system. In *Icra workshop on open source software* (Vol. 3, p. 5).

Raaijmakers, J. G. W., & Shiffrin, R. M. (2003). Models versus descriptions: real differences and language differences. *Behavioral and Brain Sciences*, *26*(6), 753–754.

Ramachandran, V. S., Blakeslee, S., & Sacks, O. W. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*. William Morrow New York.

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., . . . others (2009). Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. *Neurobiology of learning and memory*, *92*(2), 135–138.

Ratanaswasd, P., Gordon, S., & Dodd, W. (2005). Cognitive control for robot task execution. In *Proc. ieee int. workshop robot hum. commun. (roman)* (pp. 440–445).

Reiser, B. J., Black, J. B., & Abelson, R. P. (1985). Knowledge structures in the organization and retrieval of autobiographical memories. *Cognitive psychology*, *17*(1), 89–137.

Robinson, J. A. (1976). Sampling autobiographical memory. *Cognitive psychology*, *8*(4), 578–595.

Robinson, J. A., & Swanson, K. L. (1990). Autobiographical memory: The next phase. *Applied Cognitive Psychology*.

Robinson, J. A., & Swanson, K. L. (1993). Field and observer modes of remembering. *Memory*, *1*(3), 169–184.

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Conscious. Cog.*, *12*(4), 717–731.

Ross, D. A., Lim, J., Lin, R.-S., & Yang, M.-H. (2008). Incremental learning for robust visual tracking. *International Journal of Computer Vision*, *77*(1-3), 125–141.

Salgado, R., Bellas, F., Caamano, P., Santos-Diez, B., & Duro, R. (2012, May). A procedural long term memory for cognitive robotics. In *Proc. ieee workshop evol. adapt. intell. sys. (eais)* (pp. 57–62).

Salter, T., Dautenhahn, K., & Bockhorst, R. (2004). Robots moving out of the laboratory-detecting interaction levels and human contact in noisy school environments. In *Robot and human interactive communication, 2004. roman 2004. 13th ieee international workshop on* (pp. 563–568).

Schank, R. C. (1982). *Dynamic memory: A theory of learning in people and computers*. Cambridge: Cambridge University Press.

Schank, R. C., & Abelson, R. (1977). *Scripts, goals, plans, and understanding*. Hillsdale, NJ: Erlbaum.

Schillaci, G., Bodiroža, S., & Hafner, V. V. (2013). Evaluating the effect of saliency detection and attention manipulation in human-robot interaction. *International Journal of Social Robotics*, *5*(1), 139–152.

Schulkind, M. D., & Woldorf, G. M. (2005). Emotional organization of autobiographical memory. *Memory & Cognition*, *33*(6), 1025–1035.

Scogin, F., Welsh, D., Hanson, A., Stump, J., & Coates, A. (2005). Evidence-based psychotherapies for depression in older adults. *Clinical Psychology: Science and Practice*, *12*(3), 222–237.

Shaw, J., & Tiggemann, M. (2004). Dieting and working memory: preoccupying cognitions and the role of the articulatory control process. *British Journal of Health Psychology*, *9*(2), 175–185.

Shih, R., Dubrowski, A., & Carnahan, H. (2009). Evidence for haptic memory. In *Eurohaptics conference, 2009 and symposium on haptic interfaces for virtual environment and teleoperator systems. world haptics 2009. third joint* (pp. 145–149).

Simons, J. S., Koutstaal, W., Prince, S., Wagner, A. D., & Schacter, D. L. (2003). Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. *Neuroimage*, *19*(3), 613–626.

Smith, D. A., & Graesser, A. C. (1981). Memory for actions in scripted activities as a function of typicality, retention interval, and retrieval task. *Memory & Cognition*, *9*(6), 550–559.

Smith, D. M., & Mizumori, S. J. (2006). Hippocampal place cells, context, and episodic memory. *Hippocampus*, *16*(9), 716–729.

Smith, E. E., & Kosslyn, S. M. (2009). *Cognitive psychology: mind and brain*. Upper Saddle River, NJ, USA: Pearson Prentice Hall.

Smith, L. B., & Samuelson, L. K. (2009). *Objects in space and mind: From reaching to words* (The Spatial Foundations of Language and Cognition ed.). Oxford University Press.

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 101–117.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, *74*(11), 1.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *5*(1), 19–31.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory*, *82*(3), 171–177.

Squire, L. R., & Kandel, E. R. (2000). *Memory: From mind to molecules*. New York, NY: Henry Hold and Company.

Stachowicz, D., & Kruijff, G. (2012). Episodic-like memory for cognitive robots. *IEEE Trans. Auton. Mental Develop.*, *4*(1), 1–16.

Sternberg, R. J. (2003). *Cognitive theory* (3rd ed.). Belmont, CA: Thomson Wadsworth.

Stoytchev, A. (2003). Computational model for an extendable robot body schema.

Stoytchev, A. (2009). Some basic principles of developmental robotics. *IEEE Trans. Auton. Mental Develop.*, *1*(2), 122–130.

Sun, R. (2006). The clarion cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction*, 79–99.

Sutton, R. S. (2001, November). *Verification, The Key to AI.* Retrieved 2015-10-03, from `https://webdocs.cs.ualberta.ca/~sutton/IncIdeas/KeytoAI.html`

Tecuci, D. G., & Porter, B. W. (2007). *A generic memory module for events* (Vol. 68) (No. 09).

Thompson, R. F., & Spencer, W. A. (1966). Habituation: a model phenomenon for the study of neuronal substrates of behavior. *Psychological review*, *73*(1), 16.

Tielman, M., Neerincx, M., Meyer, J.-J., & Looije, R. (2014). Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction* (pp. 407–414).

Tomasino, B., Werner, C. J., Weiss, P. H., & Fink, G. R. (2007). Stimulus properties matter more than perspective: An fmri study of mental imagery and silent reading of action phrases. *NeuroImage*, *36, Supplement 2*, T128 - T141.

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological review*, *69*(4), 344.

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *381*(e402), 4.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12.

Tulving, E. (2001). Episodic memory and common sense: how far apart? *Philos. Trans. R. Soc. Lond.*, *356*(1413), 1505–1515.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.*, *53*(1), 1–25.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92.

Weng, J. (2004). Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, *1*(02), 199–236.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychological bulletin*, *121*(3), 331–354.

Williams, H. L., Conway, M. A., & Cohen, G. (2008). Autobiographical memory. In G. Cohen & M. A. Conway (Eds.), *Memory in the real world.* London: Psychology Press.

Wood, R., Baxter, P., & Belpaeme, T. (2012). A review of long-term memory in natural and synthetic systems. *Adaptive Behavior*, *20*(2), 81–103.

Woods, B., Spector, A., Jones, C., Orrell, M., & Davies, S. (2005). Reminiscence therapy for dementia. *The Cochrane Database of Systematic Reviews*, *2*.

Yildiz, I. B., Jaeger, H., & Kiebel, S. J. (2012). Re-visiting the echo state property. *Neural Networks*, *35*, 1–9.

Youssefi, S., Denei, S., Mastrogiovanni, F., & Cannata, G. (2015). A real-time data acquisition and processing framework for large-scale robot skin. *Robotics and Autonomous Systems*, *68*, 86–103.