UNSW SYDNEY

A "Society of Mind" agent architecture. A case study to modelling human behaviour in road traffic

Author:
Leu, George

Publication Date:
2013

DOI:
https://doi.org/10.26190/unsworks/16549

License:
https://creativecommons.org/licenses/by-nc-nd/3.0/au/
Link to license to see what you are allowed to do with this resource.

Downloaded from http://hdl.handle.net/1959.4/53116 in https://unsworks.unsw.edu.au on 2024-05-01

A "Society of Mind" agent architecture. A case study to modelling human behaviour in road traffic

George Leu

B. Eng. Hon. (Electronics and Communication Systems Engineering), Military Technical Academy, Bucharest, Romania

M. Eng. (Engineering and IT), National Defence Academy, Yokosuka, Japan

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy at the

School of Engineering and Information Technology

University of New South Wales, Canberra

@ the Australian Defence Force Academy

Copyright 2013 by George Leu

Abstract

This thesis proposes a non-hierarchical hybrid cognitive agent architecture, as an alternative to largely used hierarchically layered architectures which assume a predefined hierarchy of cognitive processes and, consequently, a fixed hierarchical control framework. As opposite to hierarchical view, the non-hierarchical approach considers that all cognitive processes coexist at the same level, but carry different weights in different environmental contexts; hence, a dynamic control framework.

The non-hierarchical agent architecture is based on "Society of Mind" approach on human mind and cognition, and is intended to demonstrate that such an approach can be successfully used for modelling complex cognitive processes and a wide range of behaviours.

In order to demonstrate its viability, the proposed architecture is evaluated in a cognitively demanding environment as a driver agent in traffic behaviour context. The thesis demonstrates that the resultant SoM driver agent can be successfully used as an investigation tool for a variety of road transportation issues. First, it is used in an individual setup for modelling microscopic behaviour-enabled car-following situations. Then, it is used in a multi-agent setup for modelling populations of drivers with various driving habits in order to understand how their behavioural pattern influences the traffic performance. Also, in a wider view, the thesis demonstrates that behavioural capabilities of the SoM driver agent make possible the investigation of more general transportation issues, such as the influence of human behaviour on resilience of transport systems in certain geographical areas. In a field in which cognitive-affective and behavioural issues are still treated separately from the infrastructure related aspects, a tool that is able to treat them simultaneously offers a great deal of possible benefits to the field. This thesis demonstrates that a driver agent of SoM type can become such a tool by providing the requested capabilities.

Keywords

Hybrid non-Hierarchical Agent Architecture, Society of Mind, Competing Agencies, Driver Agent, Cognitive-Affective Processes, Emotions and Personality, Road Traffic Psychology and Behaviour, Artificial Driver Agent, Road Traffic Performance, Land Transportation Resilience, City Resilience Index.

Acknowledgement

No words of gratitude are enough to thank my supervisor Prof. Hussein A. Abbass, whose untiring support, motivation, supervision and guidance in all aspects of my work provided me great help on various topics.

I also wish to thank my co-supervisor – Dr. Neville J. Curtis from DSTO – who despite being far away, amazingly, was there for me exactly in the moments when I needed most.

I owe my sincere thanks to my colleagues from the Computational Intelligence, Modelling, and Cognition Lab (CIMaC), with whom I shared wonderful moments: Dr. Sameer Alam, Dr. Jiangjun Tang, Dr. Van Viet Pham, Dr. Shir Li Wang, Dr. Amr Ghoneim, and Dr. Wenjing Zhao. I am also grateful to my dear colleagues and friends: Dr. Theam Foo Ng, Ms. Kun Wang and Mr. Nizami Jafarov.

I would like to express my appreciation for the anonymous reviewers for my journal and conference papers whose valuable suggestions helped me improve the outcome of my research.

I would like to acknowledge, with gratitude and appreciation, the financial support from the University of New South Wales Canberra @ the Australian Defence Force Academy.

Last, but not least I would like to thank to all those whom, relatives or not, I consider part of my family, for being “somewhere there” by my side through this endeavour.

George Leu

Certificate of Originality

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues, with whom I have worked at UNSW or elsewhere, during my candidature, is fully acknowledged.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

George Leu

Signed …………………….

Date ……………………......

Table of Contents

List of Figures

List of Tables

List of publications

Journal papers

- Leu, G., Curtis, N. J., & Abbass, H. (2013). A "*Society of Mind*" hybrid agent architecture and implementation of SoM driver agents for evaluation of road traffic performance and safety. Sage, (in press).
- Leu, G., Curtis, N. J., & Abbass, H. (2013). An agent-based approach for modelling deviant road traffic behaviour: artificial drivers with personality and emotions. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 17, No. 6, Fuji Press.
- Leu, G., Curtis, N. J., & Abbass, H. (2012). Modeling and Evolving Human Behaviors and Emotions in Road Traffic Networks. *Procedia - Social and Behavioral Sciences, 54*, 999-1009, Elsevier.

Conference papers

- Leu, G., Hussein, A., & Curtis, N. (2010). Resilience of ground transportation networks: a case study on Melbourne. In *The 33rd Australasian Transport Research Forum - ATRF2010*. Canberra, Australia: PATREC - Australian Planning and Transport Research Centre.
- Leu, G., Hussein, A., & Curtis, N. (2010). A GIS based assessment framework for investigating the service level resilience of ground transportation systems. In *The 31st Conference of the Australian Institutes of Transport Research – CAITR2010*. Canberra, Australia.

Chapter 1. Introduction

1.1. Overview

Agent paradigm – with concepts like agent-based systems, agent-based modelling/simulation or agent architecture – is nowadays a virtually indispensable tool for investigating a huge variety of phenomena. Use of agents allows the creation of highly versatile bottom-up tools, which can deal with complexity in both empirical and theoretical investigations.

The agent paradigm – as opposite to conventional modelling and simulation tools – became popular as an AI matter, with modelling of human decision-making and interaction. Various human stakeholders and their interaction have been modelled in order to observe or predict their individual or collective behaviour over time. Since then agent usage has rapidly spread to numerous fields of activity: from intelligent routing in telecommunication networks to modelling production plants, transport systems or financial transactions. The virtually infinite number and types of applications generated numerous agent architectures with agent designs treating with more and more accuracy various problems of interest. However, this problem-oriented approach also generated an important drawback of agent paradigm, i.e. agents are designed for solving specific problems. They are from this point of view unique and cannot be easily ported to other problems or fields of activity without significant alteration of the original design.

Indeed, formal classification of agent designs exists and any newly proposed agent architecture is referred to as “belonging” to one of the formal architectural approaches. Yet, that does not necessarily eliminate the fact that the new agent design is still unique and meant for a certain problem or set of problems of interest. Thus, the existing architectures themselves are not universally usable with any problem. Each of them covers a certain part of the whole range of complex human cognitive capabilities and can be used for a limited amount of agent-based applications. Despite agents were categorised over time in numerous classes and types, the main architectural approaches from a cognitive perspective are reactive, rational (deliberative) and affective agents. Reactive agents can actuate simple response – change in position, velocity, orientation – to local and most of the time primary stimuli such as vision or tactile perception. In contrast, deliberative and affective agents possess internal representations of the environment and are able to generate action plans and intentions based on either rational or affective decision-making processes. Since each approach is concerned with a different set of cognitive processes, the resultant agent designs are also limited to solving only the problems situated in the corresponding range.

Efforts to combine architectures into hybrid designs exit as well. They intend to cover a wider range of cognitive processes in the same time, or even the whole range of decisional capabilities of humans. A common view on building such hybrid architectures is to consider hierarchical designs. These designs usually consider that reactive processes situated at lower levels of cognitive capabilities are controlled by deliberative/affective parts situated towards the higher levels. However, resultant designs proved to become logically and computationally complex when the range of cognitive capabilities widens. An agent architecture that can cover the whole complexity of human cognitive and decision-making processes is not yet in place and efforts to create such designs are still far from showing significant improvements.

This thesis proposes an alternate architectural approach for designing complex cognitive agents. The proposed SoM architecture is presented as an alternative rather than an explicit improvement. Justification stays in the fact that the author of this thesis offers to the reader the possibility to explore untouched directions in agent paradigm. Hence, the author does not stress on the need of imposing the proposed approach to the detriment of the existing ones, but rather on the importance of opening doors towards broader views on agent design. This approach allows the design of agents with all reactive, rational and affective capabilities, outside the hierarchical paradigm. The core of a non-hierarchical view on agents is that reactive, rational and affective components coexist at the same level of cognitive capabilities, but they carry different weights in different environmental contexts. They participate in the over-all decision-making process by trying to impose their own decision in a given situation. In other words, their importance rises and falls according to the instantaneous context the agent is experiencing at a certain moment in time. This approach was introduced by Marvin Minsky in 1980's in his "Society of Mind" metaphoric book on human mind and cognition (Minsky, 1985). In Minsky's approach human mind is not a singular entity organised in a fixed hierarchical construction of cognitive capabilities. Instead, it is described as a system of competing agencies, which behaves like a dynamic, permanently evolving society-like edifice. Such a view, arguably, could describe the human nature in a more comprehensive manner than other concepts. Hence, an agent design starting from this idea would potentially very good representation of the variety of cognitive capabilities in a slender architectural framework. Thus, the thesis presents a "*Society of Mind*" (SoM) agent architecture and shows how SoM agents allow a tremendous power of representation of human heterogeneity, behavioural patterns and variety of actions.

However, the attempt to establish an agent architecture of this type also comes with the challenge of evaluating it in a cognitively demanding context. This context should cover a human cognitive-behavioural range much wider than the classical AI games or various robotics related applications. Ideally, the resultant agent design should offer support for representing/understanding behaviour of the modelled entity in an individual setup. Thus, it should deal with classic AI requests for autonomous cognitive agents such as embodiment, situatedness or coherent internal dynamics. In the same time, it should be also usable in microscopic multi-agent contexts for modelling/understanding agent interactions, relations with the environment, and emergence of collective behaviour. Last, but not least, agents of this type should be able to explain larger, general issues involving masses of agents and provide support for macroscopic studies of socio-technical phenomena.

In order to respond to these requirements the resultant agent architecture is evaluated in one of the most cognitively demanding environments, the road traffic, using a driver agent implementation. The assumption is that an SoM driver agent can be successfully used for steering the existing transportation research approaches towards a more behaviour-based perspective. Thus, the resultant driver agent is first used in an individual setup for modelling microscopic behaviour-enabled car-following situations. Then, it is used in a multi-agent setup for simulating entire driver populations in order to understand how behaviour of the population as a whole influences the traffic performance. Also, the behavioural capabilities of SoM driver agents enable investigation of more general transportation issues, such as assessing the influence of human behaviour on resilience of city-scale transport systems. In a field in which cognitive-affective and behavioural issues are treated separately from the infrastructure related aspects, a tool that is able to treat them simultaneously becomes very valuable. This thesis shows that a driver agent of SoM type can become such a tool by providing the requested capabilities.

1.2. Research motivation

Most of the existing cognitive theories and their representation in the agent paradigm as agent architectures are following the hierarchical approach. Some of these theories come with rigorous scientific arguments, while some are described in a more narrative manner. The motivation of this thesis stays essentially in the need of an alternative to the existing hierarchically layered complex cognitive architectures. The thesis starts from the assumption that various cognitive capabilities should not be seen as in a hierarchy of cognitive skills, with lower and higher levels. Minsky's approach on human mind and cognition is built on such a non-hierarchical assumption, and it created controversy at its time upon the lack of scientific validation of its claims. Many of these claims were of an unacceptable simplicity, being somewhat apart from the existing trends in both the classic AI and the newly emerging cognitive science. However, Minsky's proposal remained since then a metaphor on human mind rather than a cognitive theory. The literature available at the moment shows that it has not been addressed by studies in agent paradigm, and also no agent architecture has ever been built around the core of Minsky's vision.

Thus, the thesis attempts to identify the key elements of Minksy's approach and to use them in an agent architecture that offers high power of representation but also preserves a slender and logically simple design.

Arguably, such an agent design can be successfully used in investigation of road traffic performance, a domain in which behavioural aspects generated by human cognitive-affective processes are still insufficient studied. Indeed, most of the transportation literature formally treats behavioural issues separately from infrastructure-related issues. In real life though, they are inseparable and strongly dependent on each-other.

The SoM architecture is instantiated as a driver agent assuming that resultant implementation is capable of representing the variety of human driver behaviours. This agent could become a valuable tool for investigating and understanding the influence of human driver behaviour on the overall road traffic performance. By performing various actions coming from decisions that are sometimes rational and sometimes emotional, drivers contribute to the alteration of system performance. If the way drivers think and act is not known or taken into account at initial stages of the infrastructure planning, then planning outcomes may be inconsistent with the real behaviour of the system.

This generates the need to rethink the accustomed understanding of what planning and performance assessment of road traffic networks should do in order to create efficient road transport systems. The usual approach is, if in a certain area the reported traffic quality is below an accepted level, that is nevertheless because of intrinsic infrastructure limitations. The flow of vehicles on the roads and the way the individual drivers participate with their decisions and actions to the overall quality of traffic are not usually questioned. Vehicles and vehicle flow are only treated in an idealised manner as in other types of networked infrastructures – i.e. flow of water/gas through pipes or flow of electricity through electrical cables. Thus, planners propose various infrastructure changes with the purpose of increasing road capacity, connectivity and other infrastructure-related elements. Changes are then tested in simulated environments that use non-behavioural idealised traffic models and eventually implemented in real world. However, cases are when low quality of traffic is not because the infrastructure cannot accommodate the existing number of vehicles. Many times drivers behave in such ways that roads are always congested even if they are well designed, well built and well maintained. If these behavioural aspects are included in the assessment methodologies and models then, arguably, assessment outcomes should be more reliable and more useful for planning activities.

This thesis creates the premises for such a shift in research of transportation performance based on the proposed SoM agent architecture and its implementation as a SoM driver agent.

1.3. Research questions

In light of the facts and ideas that motivate this research, the main goal of this thesis is to create an architecture for cognitive agents which is outside the hierarchically layered approach generally used for designing complex cognitive agents. Along with this come other aspects that need to be treated, such as: finding a pertinent context in which the resultant design can be used, creating particular agents for that context, evaluating the resultant agents in that context, etc. The aspects above can be summarised in one primary research question, which further generates several important subsequent questions.

Main research question:

What is an appropriate non-hierarchical architectural approach that allows for simultaneous representation of mixed cognitive capabilities and behavioural patterns?

Subsequent questions:

1. What is an appropriate cognitive theory to support such an architectural approach?
2. How can the architecture be validated in a relevant environment?
3. Can the architecture be successfully used in real-world practical contexts?

1.4. Original contribution

The main contribution of the thesis is as follows:

- Establishes an alternative point of view on design of complex cognitive agents. It identifies Minsky's "Society of Mind" metaphor on human mind and cognition as an appropriate starting point for generating a non-hierarchical hybrid agent architecture.
- Applies and evaluates the architecture in a context that is relevant for the wide-range of cognitive-behavioural capabilities assumed in the architectural design. The architecture is instantiated as a driver agent and further tested in a road traffic context.
- Evaluates the usability of the architecture in a practical context. The Society of Mind driver agents are used in a case study for exploring the broad implications of human behaviour on the resilience of real-world road transport systems.

1.5. Organisation of the thesis

The thesis is organised in seven chapters, as follows:

Chapter 1, the introduction, presents an overview of the research field, the motivation for this study, and the research questions arising from the debated problems. The chapter also provides an outline of the thesis and the scientific contributions stemming from this research.

Chapter 2 provides a review of the literature in two major directions that are essential for this study. First, the literature on agent paradigm is thoroughly investigated, with emphasis on cognitive agent architectures and the cognitive theories which generated them. Second, an in-depth literature review is also performed on the main cognitive-affective processes that generate human behaviour, emphasising on those with high impact on traffic psychology and behaviour.

Chapter 3 presents Minsky's view on human mind and cognition, provides a thorough discussion focused on the main elements, and introduces the general

Society of Mind agent architecture. It describes the internal dynamics of the SoM agent and proposes the general methodology – as well as the formalism – for implementation of such an agent. (The chapter addresses research question 1.)

In **Chapter 4** the general SoM agent architecture proposed in Chapter 3 is instantiated in a cognitively demanding context: the road traffic psychology and behaviour, as a car driver agent. The SoM driver agent is tested in various traffic situations in order to demonstrate that such an implementation can produce a wide range of driving behaviours. The chapter describes the particular elements of the implementation of SoM driver and then evaluates its behaviour in an individual setup, in a car-following context. Evaluation is done by recording its internal dynamics and discussing the consistency of its internal decision making mechanism and the effect it actuates in traffic conditions. (The chapter addresses research question 2.)

Chapter 5 presents the SoM driver agent in a multi-agent setup from a mixed agent-transportation perspective. It shows how the *Society of Mind* paradigm and the resultant agent architecture can become useful tools for investigating behavioural aspects involved in system-level performance of road traffic networks. The chapter mainly investigates the influence of populations of drivers with different behavioural patterns on various traffic performance measures. However, it also keeps the focus on evaluating the dynamics of individual agents and the consistency of emergent collective behaviour with the individual behaviour. (The chapter continues addressing research question 2.)

Chapter 6 brings together both technological and behavioural aspects involved in the resilience of transport systems. It emphasises on the crucial role of cognitive-affective processes in formation of decision-making and emergence of human actions. Thus, it is suggested that human behaviour introduces risk in the transportation landscape in at least the same extent as the technical failures do. The investigation is performed as a case study on real road transport maps of city of Melbourne. First, it presents an investigation of the infrastructure resilience, by evaluating resilience metrics related to physical connectivity of roads and service quality of public transport. Then, an update of the infrastructure assessment is performed by adding a behavioural component to previously used infrastructure-only metrics and scenarios. The behavioural component is based on the SoM driver agent implemented and evaluated in the previous chapters of the thesis. The investigation shows how the sole use of infrastructure related resilience metrics creates an incomplete picture of city transport resilience. In contrast, the addition of human behaviour in the assessment allows deeper understanding of resilience issue and potentially more efficient planning. (The chapter addresses research question 3.)

Chapter 7 summarises the main findings of this thesis. The chapter concludes on the contributions and results, and discusses possible future research directions.

Chapter 2. Literature review

As explained in the introductory chapter the aim of this thesis is to create a non-hierarchical hybrid cognitive agent architecture starting from Minsky's "Society of Mind" metaphor on human mind. This architecture will be then used in a road traffic psychology and behaviour context for assessing the influence of human drivers on road traffic performance. In order to do so, insights on the existing advances in both agent theory and driving behaviour are necessary. Hence, this chapter is organised in three main sections.

First section presents the most important agent architectural designs and highlights their advantages and disadvantages, as well as their usage and existing implementations.

Second section presents the most important cognitive-affective aspects involved in the formation of human behaviour and discusses their particular application on driving behaviour.

Third section discusses the gaps existing in the literature in both directions and concludes on the possible ways of mitigating or eliminating them.

2.1. Related work on agents

Agents have been defined in many ways over the years and have been categorised in many taxonomies based on three major concepts of agent paradigm (Russell & Norvig, 1995; Salamon, 2011): type of environment in which the agent acts, type of interactions between agents, and agent's internal architecture. A similar categorisation was made by Pfeifer and Scheier (Pfeifer & Scheier, 1999) in their investigation on understanding the underpinnings of intelligence. They consider that design of agents consists of three essential aspects: the ecological niche, the desired behaviours and tasks, and the agent itself. Both views are nevertheless similar, only from a different perspective. Russell's view is closer to a classic Artificial Intelligence perspective, while Pfeifer's view is closer to a general socio-ecological understanding of intelligence.

For this reason the investigation on agent paradigm literature will concentrate in the following sections on these three constituents. In the beginning a general view on agents is provided in order to get an insight on the most important agent functions, types and purposes. Next, environments, interactions and architectures are reviewed in three different sub-sections.

2.1.1. Agents – general view

From a definition point of view most definitions, regardless the type of agent they are concerned with, place the agents in a perception-action cycle, shown in Figure 2.1. According to this view, an individual agent perceives the surrounding environment through sensors and based on that generates the appropriate course of action. Two relevant definitions supporting this view are cited below:

> "*An agent is any entity that can be viewed as perceiving its environment through sensors and acting upon its environment through effectors*" (Russell & Norvig, 1995)
>
> *"[…] agents are computational systems that inhabit some complex dynamic environment, sense and act autonomously in this environment, and by doing so, realize a set of goals or tasks for which they are designed*" (Maes, 1994)

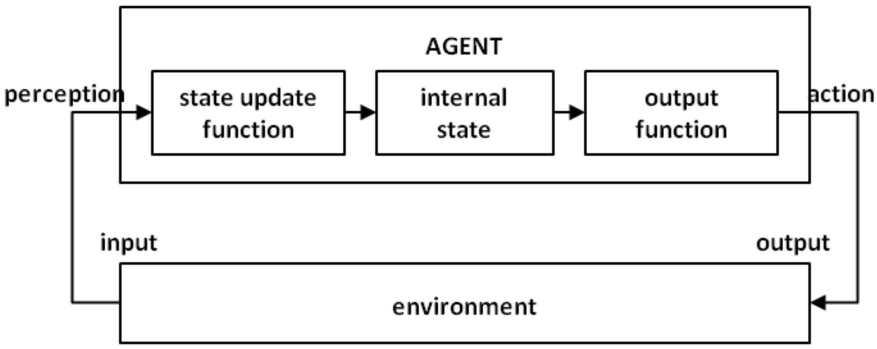


Figure 2.1 Basic view of an agent

Various other definitions have been proposed and discussed over the years (Laird, Newell, & Rosenbloom, 1987; Minsky, 1985; Ndumu & Nwana, 1997; Russell & Norvig, 1995; Wooldridge, 2002), with slight variations of meaning depending on the field of activity the authors belong to. In artificial intelligence agents must be *situated* in a certain environment in which they take decisions and actuate them according to perceived contextual changes. In distributed computing agents are autonomous software processes or execution threads. From a general modelling and simulation perspective agents represent *the model* according to which the actions and interactions of various entities occur. Also, from a socio-economic perspective agents are seen as proxies acting on behalf of other entities which cannot or do not wish to perform actions by themselves. However, since agents are used in so many fields of activity and account for so many aspects of real or artificial life, it is extremely difficult to find an over-all universally usable definition. Arguably, the definition in the Perception-Action view is perhaps the main constituent of virtually all of them and hence, the most comprehensive and generally valid.

The variety of problems for which agents are used also generated the need of agents with various properties. These properties are nevertheless problem-dependent; however, some of them are considered to be desirable in a majority of situations. Russell and Norvig (Russell & Norvig, 1995) consider *autonomy*, *reactivity*, *pro-activity* and *social ability* to be essential for building solid agent designs. Wooldridge and Jennings (Wooldridge & Jennings, 1995) also emphasise on agents' temporal continuity and ability to set goals as essential properties. Based on the above two studies and including the approach of Etzioni (Etzioni & Weld, 1995), Teahan (Teahan, 2010) presents an extended list of agent properties considered of general usability in agent paradigm. An adaptation of this list is presented in Table 2.1.

The problem-dependence and the variety of agent applications also generate difficulties in creating proper agent taxonomies with clear distinctions between the concepts used for categorisation. Due to the enormous range of applicability of agent paradigm, the boundaries between various components of certain taxonomies are rather fuzzy, with agents usually falling into multiple categories.

Agent property	Description
Autonomy	The agent exercises control over its own actions: it is able to set its own goals and to choose a way to achieve them.
Reactivity	The agent responds in a timely manner to changes in the environment.
Proactivity	The agent prepares in the best possible way to future actions that are anticipated to happen.
Social ability	The agent has the ability to communicate with other agents, including humans, in order to obtain information or help in achieving its goals.
Ability to set goals	The agent has a purpose.
Temporal continuity	The agent is a continually running process.
Mobility	The agent is able to transport itself throughout the environment.
Adaptivity	The agent has the ability to learn, and change behaviour based on its previous experience.
Benevolence	The agent performs its actions for the benefit of others.
Rationality	The agent makes rational, informed decisions.
Collaborative ability	The agent collaborates with other agents or humans to perform its tasks.
Flexibility	The agent is able to dynamically respond to the external state of the environments.
Personality	The agent has a well-defined, believable personality and emotional state.
Cognitive ability	The agent is able to explicitly reason about its own intentions or the state and plans of other agents.
Versatility	The agent is able to have multiple goals at the same time.
Veracity	The agent will not knowingly communicate false information.
Persistency	The agent will continue steadfastly in pursuit of its plan and goals.

Table 2.1 Desirable properties of agents. Adapted from (Teahan, 2010)

However, several standalone taxonomies exist. Wooldridge and Jennings (Wooldridge & Jennings, 1995) categorise agents in *weak* and *strong*. Weak agents are those entities that are capable of autonomous behaviour and possess reactive, pro-active and social skills. Strong agents are those that define in a more specific manner the concept of agency. Agents of this type have assigned specific internal states consisting of various mental elements such as beliefs, goals or intentions.

Other studies categorise agents according to their embodiment in and relation to real-life situations in natural, physical and software agents (Salamon, 2011). Natural agents are those agents reflecting equivalent natural ecosystems entities. Physical agents are the real entities themselves. Software agents are computer programs which may or may not reflect physical or natural entities. A more comprehensive taxonomy based on the same idea of agent usage and relation to real world is provided by Franklin (Franklin, 1997) and Franklin and Graesser (Franklin & Graesser, 1997). This taxonomy is shown in Figure 2.2.

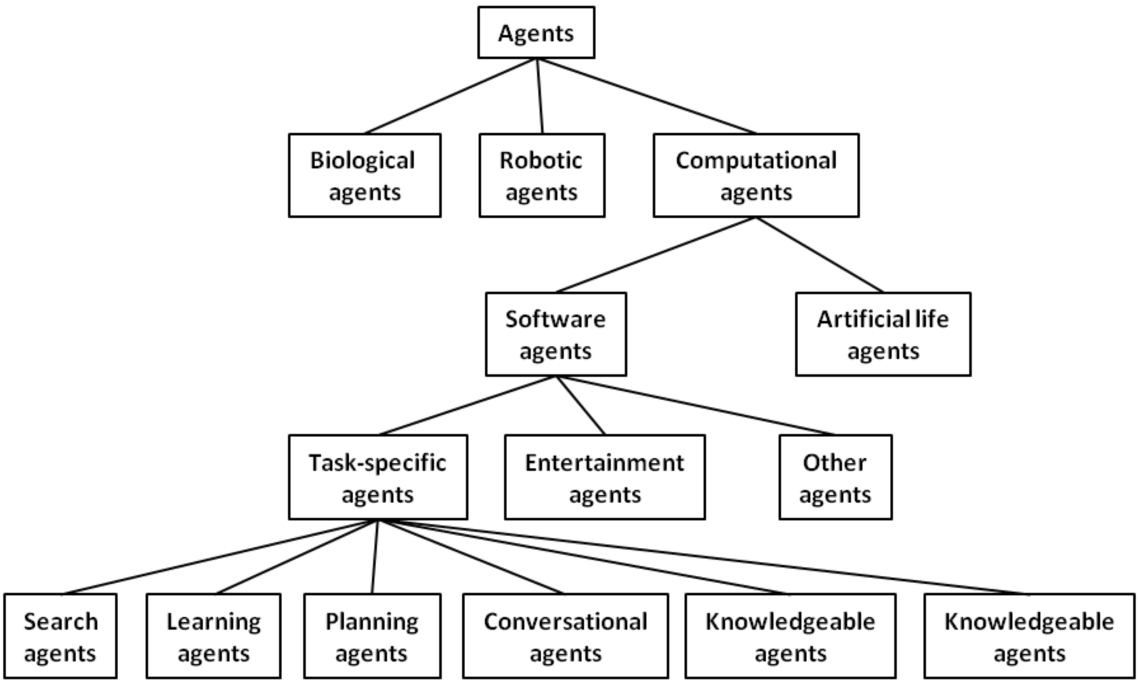


Figure 2.2 An agent taxonomy. Adapted from (Franklin, 1997) and (Teahan, 2010)

However, of the many agent taxonomies two are of major concern for this study: the perception-action cycle and the cognitive capability architectural categorisation. The former is concerned with what major steps the agent takes within the perception-action cycle. The latter shows how the agent is decomposed in a set of component modules and how these modules interact to recreate the perception-action cycle. In other words, these two approaches show WHAT and HOW the agent does for accomplishing the purpose it was designed for.

From the WHAT point of view perhaps the fundamental formulation of the perception-action cycle is the three-step SDA – Sense Decide Act – cycle (Wooldridge, 2002). In this approach data about the world are obtained through sensors (sense), then data are processed in order to establish what to do (decide) and, finally, decision is transferred into actions through effectors (act). Cycles of an SDA type involve an autonomous context-based decision making mechanism that can function based on any kind of cognitive skills, such as rational reasoning or affective processes. Different implementations of the decisional components generated numerous variations from the standard SDA cycle. Boyd's OODA – Observe Orient Decide Act – loop was used for defence/military applications (Luzwick, 2000). SIDA loop – Sense Interpret Decide Act – was used in research of organisational development as an interpretation of Sense-Response cycle (Chandi, 2005; Haeckel, 1999). Also, Sense-React cycle was used for optimisation of wireless sensor networks (Russello, Mostarda, & Dulay, 2011).

One aspect that needs to be noted is that for all the above approaches decisional stage is seen as a mono-block concept, whereas in reality decision in itself may involve several stages. This is an important aspect that needs to be addressed by the architecture proposed in this thesis, being one of issues the main research question is concerned with.

The HOW part of an agent's internal design is related to the cognitive capabilities which the respective design implements. This categorisation is actually strongly related to agent architectures and it will be thoroughly explained in the subsection that describes existing approaches on agent architectural design.

2.1.2. Environments

Environment is one of the essential components of agent paradigm. In a general view environments could be described as consisting of everything in the world created to support the purpose of the designed agent, except the agent itself and its intended behaviour. A formal definition is provided by Keil and Goldin (Keil & Goldin, 2006), who state that:

> "*An environment [...] is a physical or virtual setting that acts as the producer of the system's inputs and consumer of its outputs.*"

Environments are also dependant on the problem the agent is intended to solve. Perhaps one of the most valuable taxonomies is provided by Russell and Norvig in their comprehensive book on artificial intelligence (Russell & Norvig, 1995). According to them environments can be categorised in five criteria, or environmental dimensions.

First, an environment can be *static* or *dynamic*, with the environment being static if agent's inputs from environment are strictly dependent on its outputs to environment and dynamic otherwise. From an accessibility point of view an environment can be *accessible* or *inaccessible*, where the environment is accessible if its complete state is achievable through agent's sensors and inaccessible otherwise. An environment can be also *deterministic* or *nondeterministic*. It is deterministic if its current state and agent's current actions completely determine its next state, and nondeterministic otherwise. The authors note that in an accessible and deterministic environment agents do not experience uncertainty. If an environment is deterministic but inaccessible or partially accessible, it may appear from agent's perspective as nondeterministic. The fourth categorisation is the *episodicity*, where an environment can be *episodic* or *nonepisodic*. Episodic environments, as opposed to nonepisodic ones, are those in which agent's evolution in time is divided in episodes, each episode covering an iteration of the Perception-Action loop. During an episode, agent's action depends just on the episode itself and not on others, such as previous ones. Last criterion in the view of Russell and Norvig is *continuity*, which generates *discrete* or *continuous* environments. A discrete environment consists of a finite number of distinct components/actions from which agent can choose, whereas a continuous one consists of a range of continuous values, hence "granularity" does not exist.

A study by Keil and Goldin (Keil & Goldin, 2006) adds two more environmental dimensions to the fundamental ones of Russell and Norvig. First dimension is *persistent* versus *amnesic*. An environment is persistent if its outputs depend not only on the immediately preceding agent outputs but also on earlier agent outputs, and it is amnesic otherwise. The authors argue that persistence, from agent point of view, allows long-term expectations of reward from the environment. Second dimension is *physical* versus *virtual*, where a physical environment is defined as "*observable by analog sensors*" while and virtual one "*is accessed digitally*".

An alternate view on Russell and Norvig's *accessibility* dimension is given by Teahan (Teahan, 2010) who describes an environment as *observable* versus *nonobersvable*. However, his view is similar to *accessibility* dimension just with under a different name.

As a conclusion, even if environment taxonomies exist, clear distinctions between categories or clear inclusions in one or another category cannot be easily done. Environments still fall, as well as the agents themselves, into several categories in the same time. Environments are in an intimate relation with the agents that populate them, and hence they need to be designed for each-other in order to accomplish the purpose of modelling process.

2.1.3. Interactions

Along with the environment, agent interactions are also part of the defining attributes of agent paradigm. Interactions are mainly seen from a multi-agent perspective, describing the way agents situated in an environment influence and are influenced by others directly or indirectly in order to achieve their goals. Multi-agent systems offer consistent support for representing complex and dynamic real-world environments, such as societies, socio-ecological or biological environments. They function essentially based on the interaction between the component individual agents.

Some authors emphasise on the importance of agent interactions to the extent of considering that "*every action of an agent is part of an interaction*" (Kubera, Mathieu, & Picault, 2011). Such a view is perhaps not an exaggeration, the crucial importance of interactions being also highlighted by many other studies (Keil & Goldin, 2006; Parunak, Brueckner, Fleischer, & Odell, 2003; Wegner, 1997).

Keil and Goldin support the view of Kubera et al. claiming that interaction exists not only in multi-agent setups, but also for agents situated in individual setups.

They consider that interaction is of two main types: *sequential* and *multi-agent*. Sequential interaction involves two participants, of which at least one is an agent. This case can be associated in their view with interaction between an agent and its environment (Figure 2.3). Multi-agent interaction is different from sequential interaction through that it involves more than two agents and can be *direct* or *indirect*. Direct interaction consists of clear and distinct messages between agents in which the recipient agent is explicitly indicated in the message. Indirect interaction involves a sequential interaction between one agent (sender) and the environment. Here, the recipient agent receives the message by observing the environmental changes produced by sender agent's actions. A graphical description of this approach can be seen in Figure 2.3. Following the same idea, Kubera et al. define interactions as "*a structured set of actions involving simultaneously two agents of the simulation or an agent and its environment*".

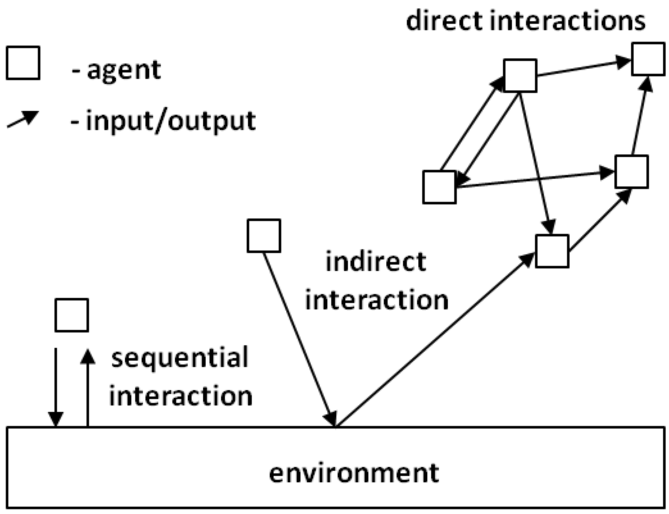


Figure 2.3 Sequential and multi-agent interactions. Adaptation from (Keil & Goldin, 2006)

Many other taxonomies for agent interactions have been proposed over time. Luck and colleagues (Luck, McBurney, Shehory, & Willmott, 2005) consider that essential for defining agent behaviour in a multi-agent context are three types of interaction: *coordination*, *negotiation*, and *communication*. Coordination assumes that agents act independently but in a coherent manner, without the need of external guidance. In their view, a particular case of coordination is *cooperation* which involves coordination with a common goal. Negotiation assumes that agents with conflicting goals interact in order to satisfy their goals simultaneously. Through this, agents gain equitable access to limited resources. In the same time, communication is viewed as the key interaction type, which facilitates all other types of interactions.

A different taxonomy is proposed by Parunak and colleagues (Parunak, et al., 2003) who consider *correlation*, *coordination*, *cooperation* and *coherence* as the main interaction types. However, unlike the approach presented in the previous paragraph communication is seen as part of coordination and not as a standalone category. On the contrary, cooperation is a standalone category, which updates the coordination by including the intentional factor. Coherence on the other hand is seen as the degree to which the pattern of agent interactions satisfies system-level goals, being rather similar to Luck's view on coordination and cooperation.

2.1.4. Agent's internal design – architectures

Previous paragraphs explained that environments and interactions define everything that is outside the agent and its relative behaviour to external context. In opposition, agent architectures are concerned with the internal structure of the agent, which generates that behaviour. Architectures establish the methodologies used for building agents, describing how the agent can be decomposed in a set of component functional blocks. They also describe how these blocks interact in order to determine the future internal state and generate actions based on the current internal state and sensorial perception. The following section discusses the fundamental general architectural approaches and explains a set of particular relevant cognitive architectures.

2.1.4.1. Reactive agents

Reactive architectures concentrate on building agents capable of fast reactions to changes detected in the immediate environment. Reactive agents have no or very simple internal representation of the environment. They are built in a behaviour-based paradigm *(Russell & Norvig, 1995)*, providing a very tight coupling between perception and action (Figure 2.4). For this reason, it is considered that reactive agents fall into a so-called Sense-React cycle. Autonomous decision-making (intelligence) is not an intrinsic attribute, but rather the product of the interaction between the agent and its environment. Existing reactive agents are mainly based on three major designs: standard stimulus-response (or condition-action) agent systems *(Milani & Poggioni, 2007)*, subsumption architectures *(Brooks, 1991)* and agent network architectures *(Maes, 1991)*.

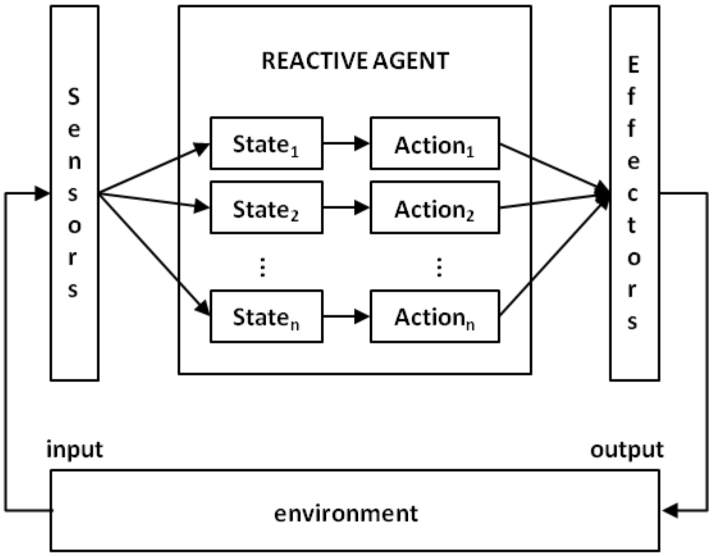


Figure 2.4 Generic reactive architecture

The tight coupling between 'Perception' and 'Action' stages in the Sense-React loop and the lack of complex decision-making mechanisms give reactive agents some inherent advantages. They are ideal in very dynamic and unpredictable environments and their implementation is simple, leading to low computational complexity. Also, without a central planning component, reactive agents have a high degree of adaptability and flexibility which makes possible their usage in massive simulations. However, their simplicity also comes with inherent limitations. Since they do not possess an internal representation of the world, decision (reaction) is based on local information only. As a result, they need sufficient information available from the local environment. Because of this reactive agents have a so-called short-term view. They do not possess long-term planning capabilities, and they are unable to take decisions based on the global state of the environment.

From a cognitive perspective, reactive agents are not intelligent per se (Russell & Norvig, 1995). In other words, they do not make use of real cognitive capabilities, but rather the multi-agent systems they are part of show an emergent collective intelligence. Hence, implementations of reactive agents are limited to simple robots or reactive entities in individual setups, and to ant colonies, swarms or bird flocks in multi-agent setups.

Though, in a different view, it can be said that some cognitive abilities exists since reactivity involves a sensorial perception (cognitive appraisal) and a simple, even though prescribed, decision making mechanism.

2.1.4.2. Deliberative agents

Deliberative architectures concentrate on long-term planning of actions (Figure 2.5) centred on a set of basic goals. They are the so-called intelligent agents, which are capable of deciding and acting in more than just a reactive manner. Decision is taken based on their own reasoning and view on the surrounding world and considering a set of alternative courses of action. The environment is represented internally as an explicit symbolic model of the world, and the decision is made through logical reasoning based on pattern matching and symbolic manipulation. Given this architectural approach, deliberative agents fall into the SDA paradigm when looked at from Perception-Action perspective.

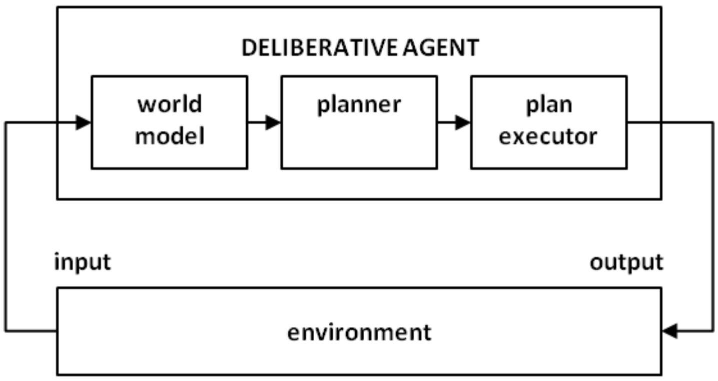


Figure 2.5. Generic deliberative/rational architecture

Advantages of deliberative agents are in that they can cover more real-world problems through the fact that decision-making mechanisms are closer to the human reasoning processes. However, deliberative agents are more computationally complex than reactive counterparts since generation of plans and processing of choice require more complex implementation and higher computational effort. Besides, they can only model rational cognitive processes and hence rational behaviour, since decision-making is based on choice and usually involves a certain profit function. The most important architectural approaches for designing deliberative agents are BDI –Believe Desire Intention – (Rao & Georgeff, 1995) and PRS – Procedural Reasoning System – (Ingrand, Georgeff, & Rao, 1992), with popular implementations such as IRMA, GRATE or PRS/dMARS.

2.1.4.3. Affective agents

In affective architectures, agents have components that in connection with other internal or external components can instantiate affective states. Thus, an affective agent contains explicit or implicit representations of various affective states (i.e. desires, emotions, moods etc.). Like deliberative agents, affective agents also fall into the same SDA cycle paradigm, but the internal decision-making mechanisms in the decisional stage are different from deliberative agents. Affective agents appeared as a response to deliberative agents inability of dealing with cognitive contexts situated outside the usability range of rational decision theory tools (Barnes & Thagard, 1996).

Appearance of affective agents was the result of a shift in AI and the emergence of cognitive science. Researchers realised that designing complex agents without affect may not be possible for the simple reason there were no agents without affect in nature at all (Scheutz & Logan, 2001), let alone the humans.

In the case of humans Damasio's (Damasio, 1994) theory of somatic markers showed that normal operation of human decision-making requires an emotional mechanism which regulates the rational reasoning. This mechanism creates biased affective “forecasts” of the potential consequences of an action. Without this emotional signal, the brain only uses rational reasoning, which slows down or even jeopardise the decision-making because of the many possible conflicting options. The assumption was demonstrated through studies on disordered patients without emotional capabilities. These patients were lost in endless rational-choice problems, being incapable of making even the simplest decision such as which outfit to wear for a certain event (Damasio, Tranel, & Damasio, 1998).

However, not only humans, but even simple organisms with no deliberative capabilities were found to have certain affective underpinnings that generate various behavioural patterns such as attraction, aversion etc. Scheuz and Logan consider that any agent of a certain cognitive complexity possesses affective states which act below or above (or both) a deliberative layer (Scheutz & Logan, 2001). A more detailed view on affective agents can be achieved from some of the most important cognitive architectural approaches such as SOAR or ACT-R which will be described in the following sections.

2.1.4.4. Complex cognitive (hybrid) agents

It is somehow clear from the above discussion that none of the main architectural approaches is suitable for building complex/complete cognitive agents. The ultimate goal of such agents is to identify and replicate, or represent computationally the mechanisms underlying human cognition. With a tremendous variety of such mechanisms studied by various sciences or fields of activity, very little can be done by trying to work on them separately. It becomes obvious that none of them can be the starting point for responding to the main research question of this thesis. Hence, the discussion continues in the following paragraphs around various ways of unifying these three fundamental cognitive processes: reactive, rational, and affective.

Perhaps the most relevant example supporting this inference is Newell's approach – the '*unified theories of cognition*' – which unifies in a single view some of the theories explaining human activity and thinking (Newell, 1990). Arguably, all existing cognitive architectures, including those reviewed later in this section, align to this approach more or less, in either a formally acknowledged way or in a de facto manner. From an architectural point of view a hybrid approach has been suggested in order to combine multiple types of cognitive capabilities and theories. The obvious and commonly used approach is to use some or all architectural approaches as subsystems in an overall complex architecture in a layered manner, as a hierarchy of cognitive capabilities. In such an architecture agent's control subsystems are arranged into a hierarchy, with higher layers dealing with information at increasing levels of abstraction. Most of the hybrid architectures consider the reactive component as representing lower level cognitive processes and providing fast response to events without complex reasoning. This is controlled by either (or both) a deliberative or an affective component situated at a higher abstraction level which contains a model of the world, and makes decisions according to rational or affective reasoning. A problem in such architectures is how to model the interactions and the control between hierarchical layers.

Two important control frameworks have been proposed by Müller and colleagues (Fischer, Müller, & Pischel, 1995), who classify such architectures into horizontally (Figure 2.6) and vertically (Figure 2.7) layered. In *horizontally layered architectures* each layer has access to sensing and acting, making possible a potential decomposition into subagents. Each layer is connected to sensorial input and action output, and so it produces suggestions as to what actions to be taken. These suggestions are "approved", "altered" or "dismissed" by the higher hierarchical layers.

A disadvantage of this approach is the informational bottleneck that can appear in the central control system. In *vertically layered architectures* sensorial input and action output are connected to the lowest layer. Hence, only the lowest layer is involved in sensing and actuating, while higher layers are involved in complex cognitive processing and decision making. A subagent decomposition becomes difficult in this case, and in addition an architecture of this type is intolerant to layer failure.

The same control frameworks were also discussed by Sloman (Sloman, 2008) in different terms. He mentions a *concurrently active* design similar to *horizontal layering* and a *pipelined* design similar to *vertical layering*. The latest is also called OMEGA design due to the resemblance of Perception-Action cycle with Greek character 'Ω' (see red arrows shape in Figure 2.7). In the same study Sloman also addresses the hierarchical aspect of layered designs. He discusses a *dominance* dimension of architectures consisting of the amount of control exercised by higher level cognitive processes onto the lower levels. The more and the stricter the control that higher levels have on lower levels, the stronger the dominance is. In the opposite direction, the dominance decreases as lower levels are allowed other types of interactions apart from the subordination ones.

The idea of hierarchy and dominance also triggers a discussion about the type of control within the hierarchy. Sloman notes that higher cognitive processes could directly turn on and off various inferior processes. Also, they can indirectly influence their operation acting as modulators, or in the less direct type of control they can facilitate their training and/or evolution.

However, most of the existing approaches have been criticised for several important aspects. Perhaps the most important one is the hierarchical approach in itself, which requires rigid control frameworks. Arguably, this limits the range of cognitive skills that designs can handle, and makes changes in design very difficult. The main research question of this thesis tries to find ways of overcoming these drawbacks, by eliminating both the hierarchy and the fixed control frameworks.

Also, fundamental sciences such as cognitive science, decision theory, AI and others have not found yet a consensus regarding which cognitive processes are positioned at which abstraction level. As a result, the existing architectures and the way they describe the interactions between various agencies (i.e. reactive, deliberative, affective) are not entirely supported by formal theories.

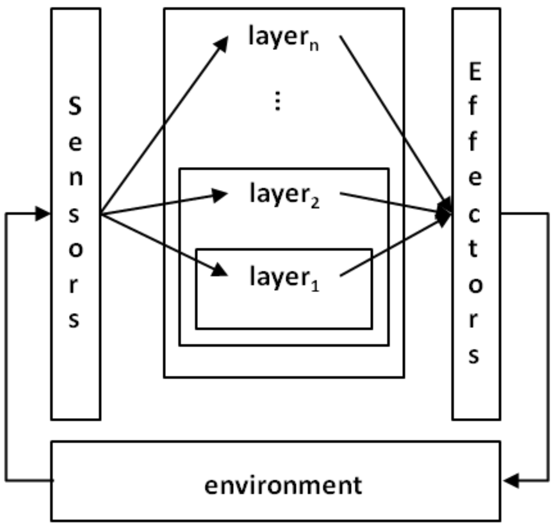


Figure 2.6 A horizontally (concurrently active) layered hybrid architectural approach

In the following paragraphs the most important and well established cognitive architectures are reviewed in order to highlight their role, scope, and advantages and limitations.

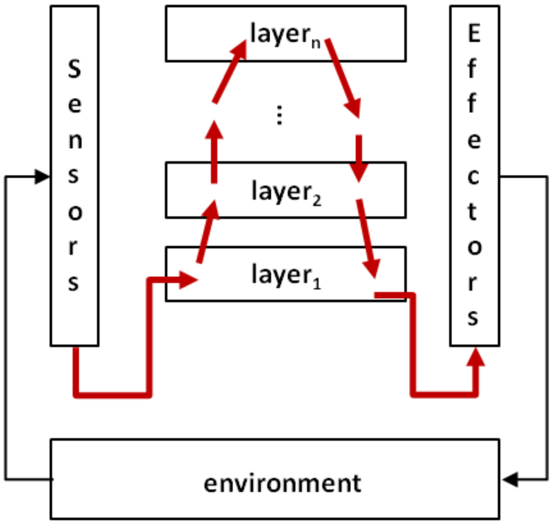


Figure 2.7 A vertically (pipelined, OMEGA) layered hybrid architectural approach

SOAR. SOAR (State, Operator And Results) is a cognitive architecture whose development started in 1980's and has been continuously improved since then (Laird, 2012; Laird, et al., 1987). It is first of all not only an architectural approach, but the expression of a complex cognitive theory that unifies a range of cognitive approaches, in way which is believed to be similar to Newell's "unified theories of cognition" (Newell, 1990). SOAR intends to offer a computational platform with the same cognitive capabilities shown by humans. It blends numerous human cognitive features such as hierarchical reasoning, learning from experience, planning, mental imagery and others. Lewis (Lewis, 2001) notes that SOAR is based on five major conceptual constituents: *physical symbol system hypothesis*, *cognitive architectures*, *production systems*, *problem spaces and least commitment control structures*, and *continuous, impasse-driven learning*. These concepts make SOAR more than just an architecture, but rather a meta-architecture of complex cognitive architectures and cognitive capabilities. As shown in Figure 2.8 and Figure 2.9, SOAR is based on two major parts. The first is a standard declarative rule-based section (*decision procedure* block) which accounts for a procedural decision-making mechanism. The second is a control knowledge section that is also rule-based and built around cognitive processes related to memory.

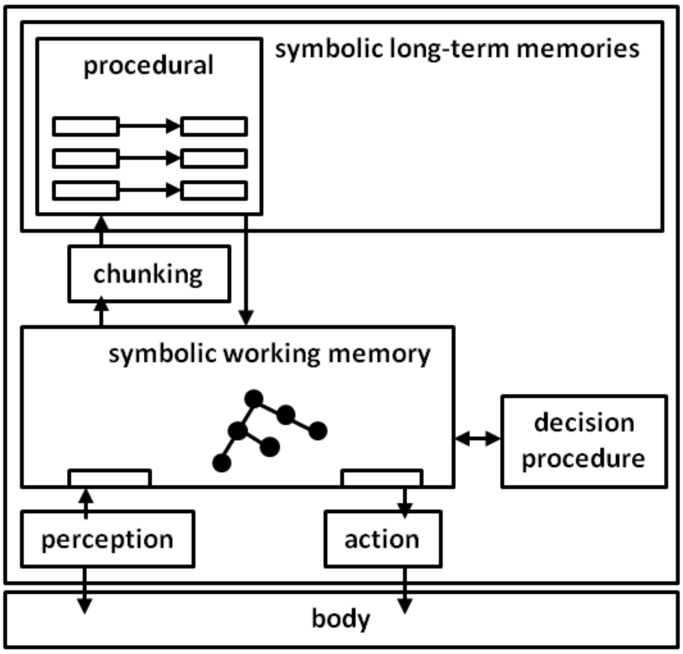


Figure 2.8 SOAR architecture – classic version. Adapted from (Laird, 2012)

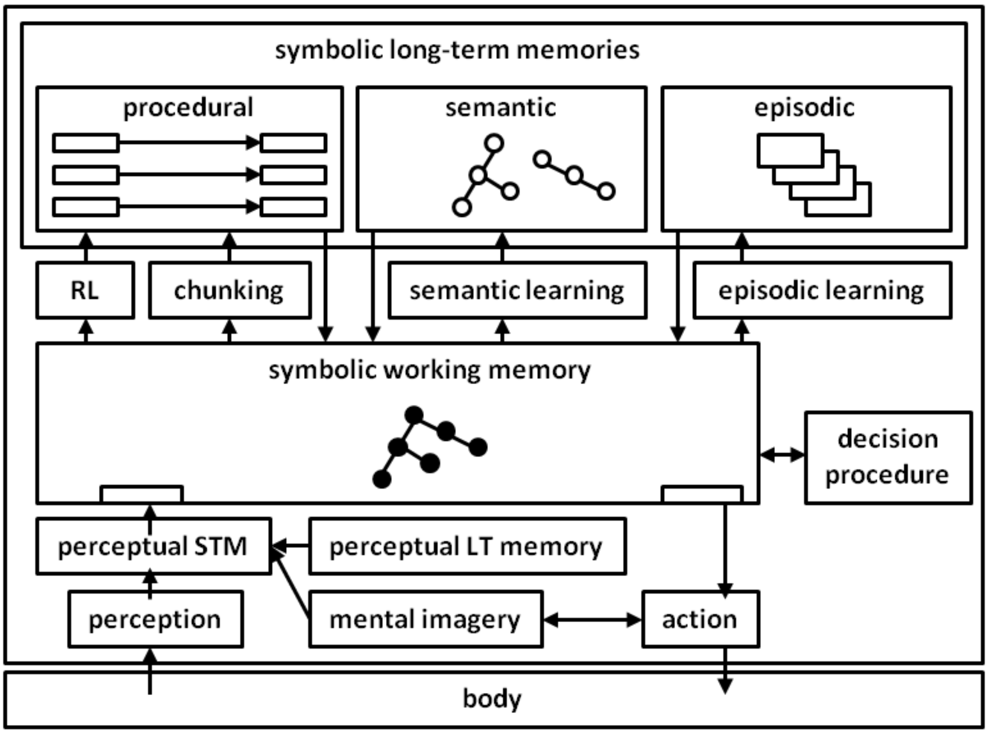


Figure 2.9 SOAR architecture – latest version. Adapted from (Laird, 2012)

In the initial versions of the architecture (Laird, 2012) symbolic long-term memory was only based on procedural memory described using “chunks” in a similar manner to that from ACT-R architecture. However, later versions gradually included *Semantic Memory* and *Episodic Memory*, allowing inclusion of more complex approaches on learning from experience such as *semantic learning* and *episodic learning*, respectively. In essence, internal dynamics of a SOAR agent can be described in an intuitive manner as a search for those cognitive operators that can take the agent closer to its goals. In this process, conflict resolution is used for decision-making in goal achievement. A critique on SOAR was that most of the architecture, conflict resolution, as well as the domain rule knowledge is described in a classic rule-based manner. Yet, this can be also an advantage that increases its portability over various applications, with only the rule base adjusted in order to perform new tasks. Also, Marinier et al. (Marinier, Laird, & Lewis, 2009) emphasise that SOAR does not use the syntax-based conflict resolution mechanisms of traditional rule-based systems. Instead, it fires all matched rules in parallel and concentrates deliberation process on the selection and application of *operators.*

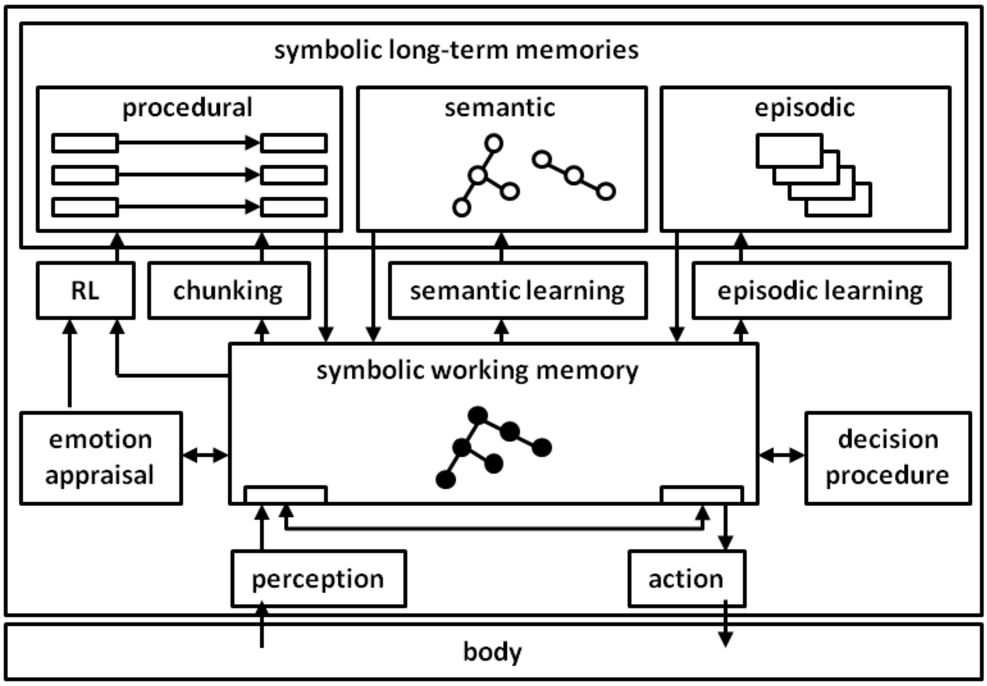


Figure 2.10 SOAR architecture with emotional appraisal. Taken from (Marinier, et al., 2009)

Lately, a highly desirable component of SOAR, which was missing during its thirty years history – the affective component, has been added to the architecture. An emotional appraisal component (*Appraisal detector* block) was explicitly added as the determinant for reinforcement learning process, as seen in Figure 2.10.

ACT-R (Adaptive Character of Thought – Rational). ACT-R is a cognitive architecture developed in the late 1990's, and improved since then, based on Anderson's ACT theory on cognition (Anderson, 1996). The architecture is mainly concerned with establishing a set of basic and potentially irreducible cognitive and perceptual/motor operations that enable the overall "operation" of a human individual.

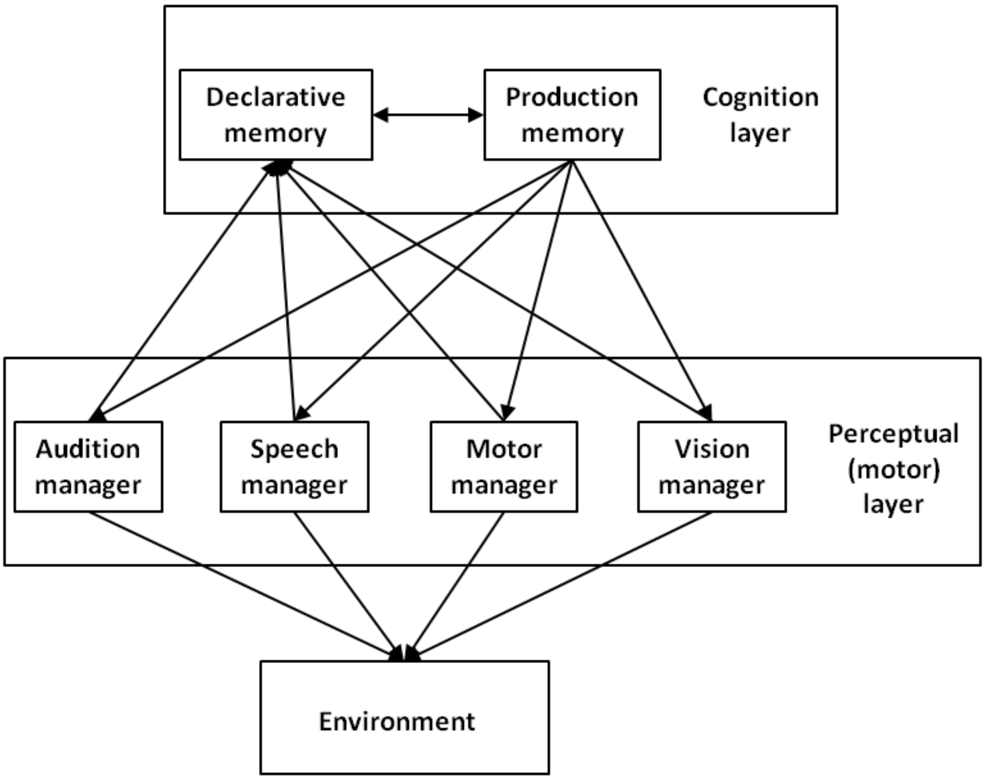


Figure 2.11 ACT-R architecture. Adapted from (Byrne & Anderson, 1997)

In ACT-R emphasis is on understanding and modelling the brain and the fundamental cognitive skills (such as vision, audition, etc.) rather than modelling the mind. On one side, this approach limits the applicability of ACT-R agents to strictly embodied and situated human agents that are able to use the human perception/motor system. On the other hand though, authors such as Stewart and West (Stewart & West, 2006) consider this approach as an important feature of ACT-R theory and architecture. They claim that simple perceptual and motor mechanisms used in ACT-R incorporate decades of research on human nature. Visual attention, audition function or other simple actions and mechanisms described in ACT-R have been solidly linked to brain regions. This provides irrefutable support and credibility to both theory and the resulting architecture. They also emphasise that the key contribution of ACT-R is the separation between declarative knowledge – '*chunks*' – and procedural knowledge – '*productions*' (see "declarative memory" and "production memory in Figure 2.11). This is also supported by a study by Salvucci (Salvucci, 2006). Declarative chunks encode simple facts, current goals and ephemeral contextual information, known as '*sub-symbolic*' parameters, which can be recalled when needed. Procedural knowledge uses production rules, usually condition-action rules, to manipulate either declarative knowledge (chunks) or the environment. Arguably, this division allows a more accurate definition of the role of control processes for various cognitive tasks.

CLARION (Connectionist Learning with Adaptive Rule Induction ON-line). CLARION is a complex cognitive architecture which was developed in the late 1990's by Sun and colleagues based on a connectionist view on cognition (Sun, 2002, 2006; Sun, Peterson, & Merrill, 1999). CLARION project was not continued with recent improvements and studies. However, it is nevertheless an important approach to agent design, which contributes to offering to the reader a wider and more comprehensive view over the agent landscape. For this reason, the author of this thesis considers that an architecture such as CLARION should not be left out of discussion in the context of hierarchical layered hybrid agent architectures.

CLARION came with a novelty element, by making a distinction between implicit and explicit cognitive processes. As a result, the architecture consists of four subsystems – action-centred, non-action-centred, motivational, and meta-cognitive – each of them having a dual representation: implicit versus explicit (Figure 2.12).

The role of action-centred subsystem is to control actions, while the non-action-centre subsystem is meant to maintain general knowledge. Motivational subsystem provides the motivation for various cognitive processes. The meta-cognitive subsystem is situated on top of the other three, monitoring, directing and adjusting their operation.

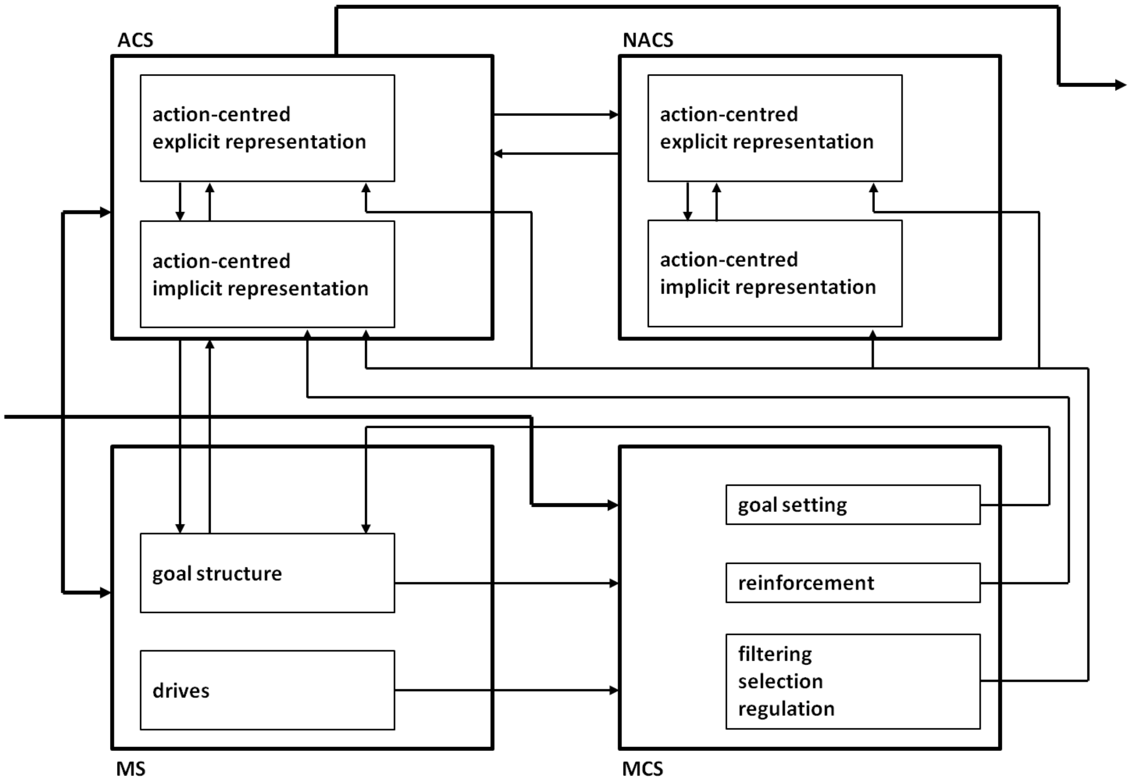


Figure 2.12 CLARION architecture: ACS – action-centred subsystem, NACS – non-action-centred subsystem, MS – motivational subsystem, MCS – meta-cognitive subsystem. Adapted from (Sun, 2006).

Throughout the years, CLARION agents have been used for solving tasks in various fields of activity, such as cognitive psychology, social psychology, intelligent systems or classic AI applications. Sun (Sun, 2006) emphasises on three aspects that make the CLARION agents capable of handling tasks from such a variety of application fields: cognition-motivation-environment interaction, bottom-up and top-down learning capabilities, and cognitive-metacognitive interaction. He also claimed that metacognitive component was a novelty element which at the time was not addressed in other cognitive architectures (However, a slightly similar approach exists in CogAff architecture which is presented later in this chapter). This element, together with the implicit-explicit representation and interactions of cognitive subsystems, allows superior skills for social interactions broadening the applicability range of CLARION agents. Also, in the same study there is a claim that the architecture would allow implicit and explicit representation and integration of innate individual biases and behavioural propensities. However, the idea is only touched, and the possibility is mentioned, but no concrete description is presented about how this was or would be done. Also, no upgrade of the architecture in this direction could be found in the either in the literature or on CLARION project's official website (http://www.cogsci.rpi.edu/~rsun/clarion-pub.html). This is one of the major issues that the main research question of this thesis is concerned with. Inclusion of behavioural biases and propensities would potentially unlock a tremendous representation power for an extremely wide range of human behavioural patterns.

CogAff (Cognition and Affect). CogAff cognitive architecture was developed mainly as part of '*Cognition and Affect*' project at University of Birmingham by Aaron Sloman and colleagues (CogAff, 2013). The project started in early 1990's and is still continuing today under various project names.

Unlike other complex cognitive architectures CogAff is presented by its authors as a "loose, informal collection of sub-projects [...] including research on architectures, forms of representation and mechanisms occurring in humans, other animals, and human-like machines". They insist that CoggAff represents more than an architecture, arguing that studying just one architecture is like doing biology by studying only one species. This view is nevertheless yielding form the same view on cognition as Newell's unified theories of cognition (Newell, 1990), from which most of the cognitive architectures were inspired. Thus, in a progress report on CogAff project Sloman avoids using the term architecture and introduces the formal architectural design using the term "schema" (Sloman, 2008). He describes the architecture as a classical hierarchical layered construction, with a reactive, a deliberative and a meta-management layer.

The design is situated within the OMEGA control framework discussed earlier in this chapter (see Figure 2.7). Yet, Sloman emphasises that CogAff agents can also act outside of OMEGA in a concurrent manner because the information flow between the components allows more than simple OMEGA representation (Figure 2.14). Also, it should be noted that the layer with highest abstraction level, the meta-management, is somewhat similar to the meta-cognitive control layer in ACT-R approach.

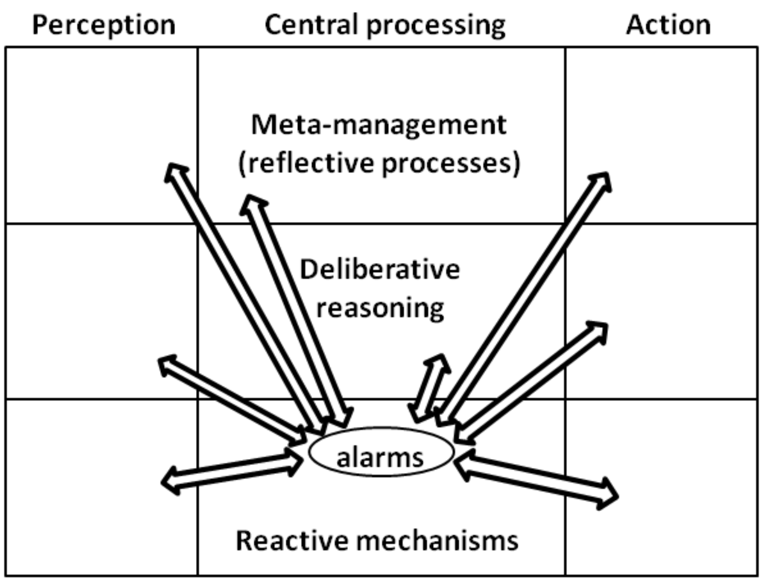


Figure 2.13 CogAff schema. Adapted from (Sloman, 2008)

However, despite being presented as a classical hierarchically layered architecture, CogAff comes with two important elements of novelty. These are presented separately in a general schema in Figure 2.13. First, *Perception* and *Action* stages are seen in a layered manner, with distinct subsets of inputs/outputs corresponding to each cognitive layer. This allows the sensorial perception system to produce specialised input for each layer of the central processing system. In a similar manner, it allows the central processing system to offer specialised response to effectors situated at various levels of cognitive response. Second, an *alarm* mechanism is considered, through which the central processing system, considered to have slow reaction, is helped to generate fast decision in rapidly variable environments. The alarm system detects urgent opportunities or danger situations which request complex calculations of the central processing system to be ruled out and replaced by fast reactions. For this reason the alarm mechanism is positioned within the reactive layer.

A drawback of this approach is that it tends to make mistakes sometimes (Sloman, 2008), due to reduction of complex reasoning processes to over-simplified reactive action courses. However the authors consider it an improvement compared to other architectures and suggest the possibility of training the alarm system as a further architectural improvement.

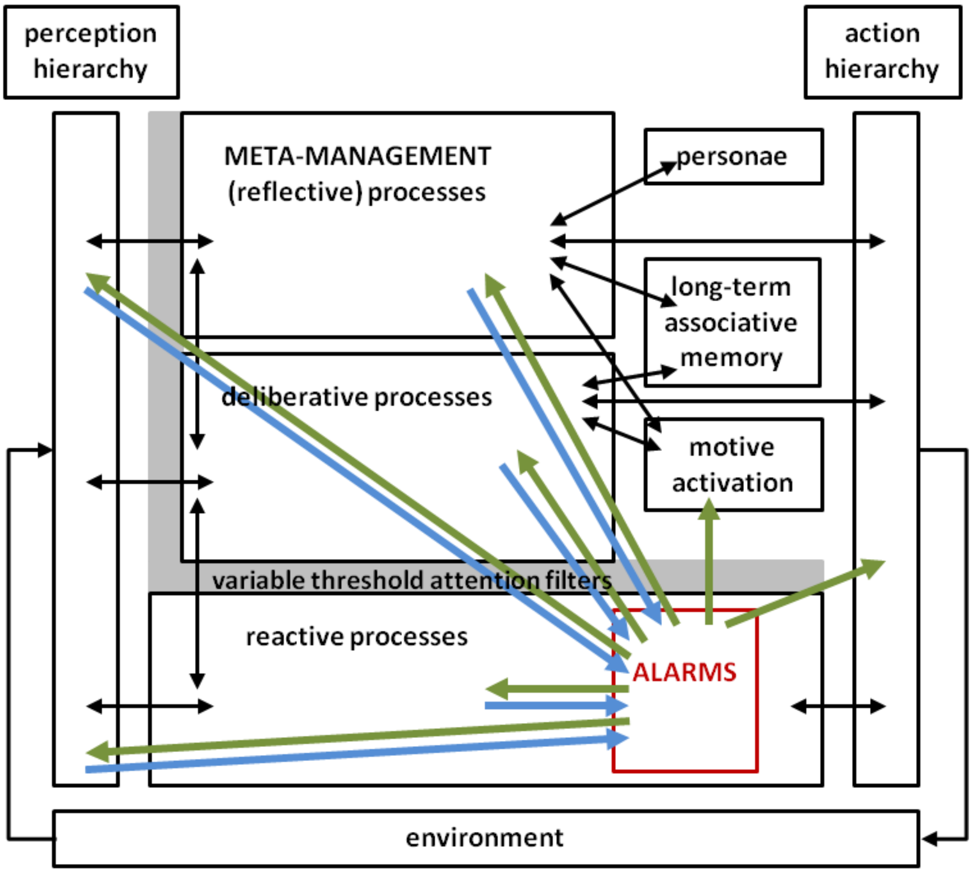


Figure 2.14 CogAff architecture. Adapted from (Sloman, 2008)

2.2. Related work on underpinnings of human drivers behaviour

Research in behavioural aspects involved in human activity, especially in driving behaviour, agree that behaviour is the resultant of two major types of cognitive features. First, there is a set of innate, unchangeable and context independent traits. Second, there is a set of features which are context-generated and variable in real-time. In particular for human drivers, as early as in the 1940's Tilmann and Hobbs noted in a very narrative manner that "*a man drives as he lives*" (Tillmann & Hobbs, 1949). In a more specific view Revelle describes the idea in terms of personality and emotions, stating that "*emotions are to personality as weather is to climate*" (Revelle, 2003).

This section concentrates on this aspect and brings into attention the most relevant findings in the fields of human personality and emotions, in order to provide a clear understanding on the factors that generate various driving behavioural patterns. This understanding creates the starting point for answering research questions 1, 2 and 3, by establishing the road traffic as a relevant context for testing the architectural assumptions made in this thesis.

2.2.1. Personality

Historically, the study of personality started as early as from the antiquity and most likely even earlier, as humans tried to find out how they are similar or different to other individuals and why. Study of personality can be viewed in two ways. In the first approach, numerous studies tried to explain the conditions in which personality emerged. The scientific investigation went from as low as the genes and chemical reactions involved in the existence of live matter to as high as social and environmental interactions. In the second approach, the researchers tried to develop descriptive taxonomies of individual differences and similarities, rather than explaining the underpinnings and formation of human personalities. These approaches use various techniques in order to establish sets of traits, features or behavioural patterns that can categorise human activity and the individual differences.

In the first approach five major fundamental theoretical directions are highlighted, for the reason of showing a complete picture of personality theory: evolutionary psychology (Figueredo, Sefcek, et al., 2005; Figueredo, Vásquez, et al., 2005), behavioural genetics (Plomin & Rende, 1991; Saudino & Plomin, 1996), biological theories (Buss, 1990; Depue & Collins, 1999), socio-cognitive theories

(Bandura, 1999; Mischel & Shoda, 1998) and psychoanalysis (Blum, 1953; Rapaport, 1960). However, the most notable findings with important impact on understanding human behaviour –driving behaviour in particular – belong to the second approach: the taxonomic one.

The taxonomic approach is grouped around the trait theories. These theories assume that individual personalities consist of broad dispositions (traits) which are stable, innate characteristics that causes individuals to behave in certain ways. Combination and interaction of these traits form the unique personality of each individual and hence the theory focuses on finding and measuring the individual differences through these personality characteristics. Trait theories start from the "lexical hypothesis" (Goldberg, 1990, 1992), according to which humans develop commonly and widely used generic terms for expressing individual differences in their daily interactions. When one is asked to describe the personality of a known person then one would provide a list of traits which is commonly used in the daily life for describing people (such as: kind, energetic, talkative, etc.).

The work of Gordon Allport (Allport & Odbert, 1936; Goldberg, 1990) can be considered as one of the first attempts of establishing a coherent taxonomy, even if important work has been done in the 19th century by Francis Galton (Goldberg, 1990). Allport catalogued 18000 English terms describing personality, and selected 4500 which were classified as stable traits (Goldberg, 1990). He suggested they can be grouped in three categories: cardinal traits, central traits and secondary traits. Cardinal traits dominate and shape a person's behaviour and based on them an individual can be strongly recognised. Central traits can be generally found in some degree in every person. Secondary traits can be observed only in certain circumstances (such as particular likes or dislikes that a only very close friend may know).

Based on the work of Allport and Odbert (Goldberg, 1990), Catell reduced the number of stable personality traits to 171, from which he eliminated closely related terms reducing the list to 16 bipolar personality factors, called primary personality factors. To each factor he associated a set of descriptors for high and low ranges (Cattell, 1965; Eysenck, 1991). Catell's work, despite being important for personality research, has been highly criticised for the fact that it could not be fully replicated (Boyle, 1989; Eysenck, 1991). However, the controversy around Catell's model and the efforts to validate it generated other models that contributed substantially to trait theory. Hans Eysenck (Eysenck, 1991, 1992) developed a three-dimensional model with three bipolar universal traits, as follows:

- **Introversion/Extraversion.** Introversion involves directing attention on inner experiences, while extraversion relates to focusing attention outward on other people and the environment.
- **Neuroticism/Emotional Stability.** Neuroticism refers to an individual's tendency to become upset or emotional, while stability refers to the tendency to remain emotionally constant.
- **Psychoticism.** Individuals with high psychoticism tend to have difficulty dealing with reality and may be antisocial, hostile, non-empathetic and manipulative.

Arguably the most important of the trait models is the five-factor model, also known as the "Big Five" (Goldberg, 1990). Big five model emerged in the second half of 20th century, starting with the work of D. W. Fiske in 1949 (Fiske, 1949) and continuing with Norman (Norman, 1963), McCrae and Costa (Robert R. McCrae & Costa, 1987), Goldberg (Goldberg, 1990) and others. The model assumes the existence of five fundamental bipolar dimensions of the human personality and places the humans in a continuous range between two extremes (low and high traits) for each of the five factors, as follows:

- **Extraversion.** This trait includes characteristics such as excitability, sociability, talkativeness, assertiveness and high amounts of emotional expressiveness.
- **Agreeableness.** This personality dimension includes attributes such as trust, altruism, kindness, affection, and other pro-social behaviours.
- **Conscientiousness.** Common features of conscientiousness dimension include high levels of thoughtfulness, with good impulse control and goal-directed behaviour. Individuals high in conscientiousness tend to be organized and mindful of details.

- **Neuroticism.** Individuals high in this trait tend to experience emotional instability, anxiety, moodiness, irritability and sadness.
- **Openness.** This trait features characteristics such as imagination and insight, and those high in this trait also tend to have a broad range of interests.

The big five model outdated other existing models in the field of lexical research and became one of the most solid approaches in the personality research. It demonstrated a remarkable universality across languages and cultures (Costa, Terracciano, & McCrae, 2001; Robert R. McCrae & Antonio Terracciano, 2005; David P. Schmitt, Realo, Voracek, & Allik, 2008), and also high capability of predicting human behaviour (Mershon & Gorsuch, 1988; Paunonen & Ashton, 2001; Saulsman & Page, 2004). Thanks to the above mentioned aspects in the last decade the big five model became an important tool in explaining, modelling or predicting human behaviour in most aspects of the real life.

Personality and traffic behaviour. In traffic behaviour, Brandau and colleagues (Brandau, Daghofer, Hofmann, & Spitzer, 2011) note that risk can be introduced in traffic as a result of two aspects. First, is because of insufficient skills, lack of experience or biased risk perception (Deery & Fildes, 1999; Gregersen, 1996; Gregersen & Bjurulf, 1996). Second, is because of the innate personality features of those involved in traffic, features that generate a certain lifestyle and behaviour. Deery and Fildes (Deery & Fildes, 1999) identified the most important types of behaviour which introduce high risk in traffic situations: aggression, competitive speed, driving to reduce tension, assaultiveness and hostility. All of them are the the result of certain personal desires, motivations and needs – sensation seeking (Jonah, 1997), novelty-seeking (Cloninger, 1994) – which are in turn generated by individual personality traits. Jonah notes that driver's personality has been considered as an underlying causal factor in traffic behaviour starting with the work of Tillman and Hobbs, who stated that "a man drives as he lives" (Tillmann & Hobbs, 1949). Evidence of a relation between sensation seeking and psychoticism dimension of Eisenck's three dimensional personality model has been found by Zuckerman (Eysenck, 1983; Zuckerman, 1994). Zuckerman believed that sensation seeking could be even treated as part of the psychoticism dimension. He also found that sensation seeking correlates moderately with impulsivity and weakly with extraversion.

Yet the main body of research investigates the involvement of the more recognised and validated Big Five model of personality in risky behaviour of all types. A meta analysis performed by Clarke and Robertson (S. Clarke & Robertson, 2005) showed that low agreeableness and conscientiousness can be considered valid predictors for risky traffic behaviour and involvement in accidents. They also found that extraversion is as well a predictor for risky behaviour, hypothesis which is supported by an earlier study of Lajunen (Lajunen, 2001).

In a different study Jovanovic and colleagues suggested that neuroticism correlates with aggressive driving indirectly through the affective state of anger, while the other traits are directly correlated (Jovanovic, Lipovac, Stanojevic, & Stanojevic, 2011). They found that three of the five personality traits – neuroticism, agreeableness and conscientiousness – were strongly correlated with risky behaviour and were appropriate for predicting driving related anger and aggression. Similar results have been also reported by Miles and Johnson (Miles & Johnson, 2003) who also found that low agreeableness, low conscientiousness and high neuroticism account for aggressive driving. In the same study, comparison of survey responses of groups with different driving experience, attitudes and believes (e.g. experienced drivers and students) revealed a fairly invariant pattern of the personality traits involved in aggressive driving among the subjects.

This suggested that personality traits, despite being related to human factors and affective states, have also a significant universal stand-alone influence on driving behaviour. The assumption was supported by large scale social studies which used national social indicators on 18 (Lynn & Hampson, 1975) and 32 (Lester, 2000) nations. Results showed similar correlation patterns between personality traits, and risky behaviour and accidents of all types among the whole set of surveyed countries. In agreement with this assumption other researchers (Classen, Nichols, McPeek, & Breiner, 2011) even argued that personality testing should be added to driving skills assessment battery.

2.2.2. Emotions

Emotions have been intensely studied in the last decades as part of numerous fields of research in an attempt to assess, quantify, predict or model the human behaviour in various real-life or virtually generated scenarios.

Lazarus (Lazarus, 1982) and Damasio (Damasio, 1994) treated the affective processes as melded with the cognitive processes in a single cognitive-affective system, suggesting that emotions involve cognitive activity. This explains the appearance of emotional states as result of the perception of different stimuli and also explains the decision-making as a result of emotional activity. In a different view, as a direct reply to the views of Lazarus (Lazarus, 1982), Murphy and Zajonc (Murphy & Zajonc, 1993) claim that emotions can be elicited by a minimal stimulus input and virtually no cognitive processing. Thus, the affective system is separated from the cognitive one. An intense argumentation between the two approaches, consisting of a long series of direct replies, took place in the "American Psychologist" over a period of several years at the beginning of 1980's (Lazarus, 1981, 1984; Zajonc, 1984). It ended with another reply to both approaches (Kleinginna & Kleinginna, 1985) which signalled that both were correct with respect to their goals, but the goals were different since the authors defined the affect and cognition in different ways.

The divergence of opinions not only took place among different researchers, but also within the work of same researchers, who oscillated over time in supporting one or the other based on the findings of the time. As an example, in the early 1990's in his book "Mind: introduction to cognitive science" Paul Thagard did not mention emotions as part of his methodology for investigating the cognitive system. Several years later, in the second edition of the same book (Thagard, 2005), he reconsidered his position concluding that emotions could and should be included in a larger system of cognitive processes. Yet, the purpose of this chapter is not to challenge the fundamental approaches on cognitive-affective processes.

It is important to say though, that emotions have been demonstrated to play a crucial role in behaviour of individuals, regardless the approach or field in which they have been studied (Dolan, 2002; Gigerenzer & Gaissmaier, 2011). In a neurological approach, Damasio shows in his "somatic markers" hypothesis that emotions are biologically indispensable to decision making and hence to human behaviour (Damasio, 1994; Damasio, et al., 1998). He shows that people with damage in brain areas hosting emotional centres (Pessoa, 2008) keep intact all their social knowledge, rational reasoning and interaction capabilities but lose their ability to form any emotional response. These patients become too rational and most of the time they are trapped in endless analyses of numerous and conflicting options. Eventually, they are unable to decide even in the simplest situations such as choosing a restaurant or the outfit for a certain social event.

In a rational approach Thagard and Millgram consider in their “coherence theory of decision” that individuals choose out of the available plans the one which is the most coherent with their current goals (Thagard & Millgram, 1995). Later, Barnes and Thagard acknowledge and support Damasio’s hypothesis and they include it in an extended version of the coherence theory called “the theory of emotional decisions“ (Barnes & Thagard, 1996). In this theory the inconsistencies between available plans and the current goals elicit emotions that facilitate decision by generating new plans which are more coherent with the existing goals. The authors state that “*emotions function to reduce and limit our reasoning, and thereby make reasoning possible*“. Their statement is also supported by the “*heuristic decision making*” approach of Gigerenzer and Gaissmaier (Gigerenzer & Gaissmaier, 2011).

From a taxonomic point of view two main directions of research emerged in the field of emotions, a discrete and a continuous dimensional approach (Mauss & Robinson, 2009).

In the discrete approach the attention is concentrated on analysing behavioural patterns in order to recognize and categorize the basic human emotions (Ekman, 1999; Izard, 1992; Ortony & Turner, 1990). In this case researchers consider that each basic emotion corresponds to an unique behavioural pattern, hence, the challenge of finding the most appropriate set of basic emotions (Panksepp, 2007). Ortony and Turner (Ortony & Turner, 1990) provide a comprehensive list of sets of basic emotion based on the most important studies in the field over a period of more than 20 years. Arguably, there is not yet such a set of standard basic emotions, but six of them are considered in the literature as the “big six” which express the main aspects of human behaviour: anger, fear, sadness, happiness, surprise and disgust (Ekman & Friesen, 1971).

Dimensional approach considers there are several fundamental emotional primitives, bipolar and continuous, which can represent the variety of human emotional responses. There is almost a consensus that the emotional space consists of three dimensions: valence – pleasure/displeasure, activation – low arousal/high arousal, and dominance – approach/avoidance (Bolls, Lang, & Potter, 2001; Grimm, Kroschel, Mower, & Narayanan, 2007; Lang, Dhillon, & Dong, 1995). Similar to the five dimensional approach on personality (Big Five model), individuals are placed for each of the emotional dimensions somewhere in a continuum between the two bipolar extremes.

Despite being different in the way they define and conceptualise the emotional states, some researchers tried to connect the two approaches. They claimed that each discrete emotion can be represented in the dimensional space as a combination of emotional primitives (Haidt & Keltner, 1999; Thagard, 2005).

Emotions and traffic behaviour. In traffic behaviour research about the risk introduced by emotions, argues around two main aspects: risk introduced by unintentional distraction from driving and risk introduced by intentional aggressive driving. In the first case, drivers unintentionally perform dangerous actions such as speeding, eating, phoning etc. as the result of inattention. In the second case actions are volunteer and they manifest through obviously hostile attitudes like volunteer tailgating and lane cutting, aggressive body gestures towards the other traffic participants etc. (Deffenbacher, Deffenbacher, Lynch, & Richards, 2003; Dula & Geller, 2003; Pêcher, Lemercier, & Cellier, 2009). While both of them translate into risky behaviour in traffic conditions, the affective states involved in the two behavioural patterns and their elicitors are nevertheless different (Pêcher, et al., 2009).

From the perspective of distraction from driving Fiedler and Bless (Fiedler & Bless, 2001) showed that sadness, as a negative emotion, narrowed the attentional focus to particular elements of the environment. This generates longer reaction time, cognitive appraisal distortion, rumination and self-focus (Gotlib & McCann, 1984; Lyubomirsky, Kasri, & Zehm, 2003; Silvia & Abele, 2002; Silvia, Eichstaedt, & Phillips, 2005). On the contrary, positive emotions such as joy have been proved to broaden the attentional focus. This significantly improves the reaction time, the situation awareness and the ability to anticipate changes in the traffic environment (Fiedler & Bless, 2001; Rowe, Hirsh, & Anderson, 2007). Though, extreme states of positive affect are considered to produce excessively high self-confidence which can result in greater risk-taking attitudes (Wright & Bower, 1992).

Grimm and colleagues (Grimm, Kroschel, Harris, et al., 2007) reached the same conclusions by using the dimensional approach on emotions. They considered happiness, anger and sadness as the main "directions" for representing driver emotional states while driving, each of them being situated in the activation-valence-dominance emotional space. On "happiness" direction lower values create the so-called "optimal flow" (slightly positive valence, moderate activation) which increases attention, focus and productivity. Higher values create the "extreme happiness" (high positive valence, high activation) which facilitate inattention and distraction from driving. On the "sadness" direction lower values (slightly negative valence, low activation) degrade the task performance and facilitate inattention. Higher values generate "drowsiness" (very low valence, very low activation) which increases even more the task degradation and inattention. Anger direction in Grimm's study accounts for the perspective of intentional aggressive driving (Javela, Mercadillo, & Ramirez, 2008; Lerner & Keltner, 2001) rather than for "distraction from driving". Lower values on this direction represent "frustration" and they can generate disrespect in traffic and unlawfulness. Higher values account for "anger" and generate aggression, road rage and risk-seeking behaviour.

For a better understanding of the relation between emotional states and intentional risky behaviour Shinar proposed a frustration-aggression model (Shinar, 1998). In this model negative emotions corresponding to "anger" direction of Grimm et *al*. determine inconsiderate or annoying acts directed to others and deliberate dangerous driving behaviour. Implication of anger in risky and illegal driving, social irresponsibility and high crash rates has been constantly reported over the years by many other studies (Deffenbacher, et al., 2003; Ellison-Potter, Bell, & Deffenbacher, 2001; Lajunen & Parker, 2001; Mayer & Treat, 1977; Sarkar, Martineau, Emami, Khatib, & Wallace, 2000). In a taxonomic view of the aggressive driving and road rage on freeways, Sarkar and colleagues (Sarkar, et al., 2000) proposed a labelling system with five categories of aggressive driving behaviour, as follows:

- “Speeding Alone” – consists of speeding only;
- “Aggressive Driving 1” – consists of speeding and unsafe lane changes and passing;
- “Aggressive Driving 2” – includes weaving and cutting through congested traffic without speeding;
- “Aggressive Driving 3” – incorporates tailgating and “Road Rage”, which accounts for intimidating attitudes such as verbal harassment, threats, obscene gesturing, malicious braking etc.

They found that the most frequent risky behaviour referred to “Aggressive Driving 2” (weaving and cutting) with a percentage of 27.1, while the least frequent was “Aggressive Driving 3” (tailgating) with 12.5 percent. In a different view, other studies related anger with speeding in adolescents and elevated state anger with reckless driving in college students (Arnett, Offer, & Fine, 1997).

2.3. Discussion

First section of this chapter showed that existing agent architectures are not, as expected, universally usable with any problem. Rather, each of them covers a certain part of the whole complex human cognitive capabilities and can be used in a limited range of agent based applications. For example, reactive agents can actuate simple response to local and most of the time primary stimuli such as vision, tactile perception etc. This response is most of the time in the physical domain such as change in position, velocity, orientation etc. As opposite to reactive agents, deliberative and affective agents possess an internal representation of the environment and are able to generate action plans and intentions based on either rational or affective decision-making processes respectively. Since each approach is concerned with a different set of cognitive capabilities, the resultant agent designs are also limited to solving those problems which are in the respective range of cognitive abilities.

Efforts to combine architectures into hybrid designs in order to cover a wider range of cognitive capabilities in the same time exist as well (Marinier, et al., 2009). A common view on building such hybrid architectures is to consider hierarchical designs with reactive parts at the lower levels of cognitive capabilities controlled by deliberative or affective parts situated at higher levels. However, resultant designs proved to become logically and computationally complex when the range of cognitive capabilities broadens, fact that limits their usability. An agent architecture that can cover the whole complexity of human cognitive and decision-making processes is not yet in place and efforts to create such designs are still far from showing significant improvements. This generates the need of addressing the main research question of this thesis.

Arguably, Minsky's view on human mind and cognition could be a starting point for building "complete" agents outside the hierarchical paradigm. The core of a non-hierarchical view on agents would stay in the fact that reactive, deliberative and affective components coexist at the same cognitive capability level, but they are active in different contexts. They participate to the over-all decision-making process by trying to impose their own decision in a given situation. In other words, their importance rises and falls according to the instantaneous context the individual is experiencing. Minsky claims that such a view describes the human nature in a more comprehensive manner than other concepts proposed before. Based on this claim, two aspects can be suggested. First, an agent architecture starting from this idea would potentially offer much better representation of the variety of cognitive capabilities. This is due to the fact a virtually unlimited number of competing agencies can be included in the design, each of them implementing a specific cognitive capability. Second, the resultant design can be simpler and less computationally complex. This is due to the fact the complex framework for layer control is actually inexistent.

However, as seen in the second section of this chapter, there is indeed a huge variety of human cognitive capabilities, and accordingly, a huge range of resultant behavioural patterns. Hence, an eventual driver agent implementation of an agent architecture based on Minsky's view should handle this variety by taking into account both major types of mental features: the real-time context generated emotional states and the non-contextual innate personality traits. These aspects will be explained in detail in the next chapters.

Chapter 3. A general SoM agent architecture

Existing agent architectures – reactive, deliberative or affective – are meant to treat limited sets of cognitive capabilities fact that makes them usable for problems situated in those particular cognitive frameworks. Various hybrid architectures were also proposed, in an attempt to combine these three major approaches for modelling complex cognitive tasks. Most of them are layered designs that place the simple architectures mentioned above in hierarchical frameworks, assuming that some cognitive processes are situated either below or above others. This approach has important drawbacks related to design complexity and computational cost. In addition, there is insufficient theoretical support. Designs are based on multiple bits of theoretical approaches, which, combined, result in designer's vision rather than in the summation of those approaches.

This chapter introduces an alternate design to those mentioned above based on Minsky's "Society of Mind" view of human mind and cognition, in an attempt to mitigate some of the drawbacks of the existing designs.

The chapter is organised in five major sections. First section discusses Minsky's "Society of Mind" approach, its advantages and disadvantages. Second section proposes a potential computational instantiation and sketches a possible general architectural design, presenting an architectural schema. Third section presents a detailed view on the proposed architecture, explains the internal dynamics of a SoM agent and describes in detail potential the methodologies for building such agents. Third section presents a general implementation methodology usable for building various particular instances of general SoM agents. Fourth section shows a JAVA class structure and presents some concrete implementation facts. Last section concludes on the contribution of such an architectural approach to research in general agent paradigm, and also on its applicability range.

3.1. Minsky's "Society of Mind" metaphor on human mind and cognition

The "*Society of Mind*" (SoM) paradigm proposed by Marvin Minsky in mid 1980's (Minsky, 1985) is largely perceived as a metaphor about the human mind, rather than as a scientific theory. However, it may contain the answer for the difficult question of modelling the interactions between various classes of cognitive processes, in order to design complex cognitive agents. SoM is a valuable vision on the human mind, on how the human actions emerge from a so-called "heterarchy" of interacting and competing internal entities. These entities can be ideas, sensorial perceptions, memory of past actions etc. or the effect of those. Some scientists from AI, cognitive science or computer science disapprove the lack of scientific validation of the many concepts and ideas presented in Minsky's approach (Ginsberg, 1991; Reeke Jr, 1991; Smoliar, 1991). However, they also acknowledge that any single claim made by SoM is true and undeniable in its common sense. Minsky's approach, despite not qualifying for a theory, it is virtually unarguable and it comes from a pertinent observation and very deep understanding of human nature and human action formation mechanisms (Dyer, 1991; Stefik & Smoliar, 1991; Thagard, 1993).

The lack of scientific validation from a narrow point of view of validating a model or a theory is replaced by validation through whole fields of activity. Many of the facts presented in Minsky's metaphor have been clinically demonstrated in fields like neuro-pathogoly, neuro-psychology and medical-pathology. They have been observed in behaviour of and studied on numerous human subjects, patients and disordered brains (Hermans, 2002). However, they have not been described in relation with Minsky's approach. Minsky's view was actually redefined many years after SoM was published, as "the dialogical self" (Hermans, 1996, 2001), an approach in which Herman sees the individual "self" as a "heterogeneous society".

Minsky's book on SoM is very wide though. It covers virtually all aspects of human mind which are of great concern in AI, cognitive science or computer science, just from an alternate perspective (Stefik & Smoliar, 1991). Despite being received with a certain reserve by the scientific world, his book triggered a vivid interest in creating computational models and architectures of mind. However his approach was never instantiated as it was in a computational model/ architecture. It remained only a narrative endeavour, even though some agent architectures implemented here and there parts of Minsky's view.

Among the many concepts and assumptions Minsky's theory is built on, several aspects are fundamental for this thesis: the principle of non-compromise, the k-line theory of memory (where memory is not treated as in the literal/technical meaning, but rather in a wider SoM view) and the agent heterarchy. These concepts will be discussed in the following sections in order to establish their role and scope both the Society of Mind context and the resultant SoM agent architecture.

3.1.1. Conflict and compromise – the principle of non-compromise

According to Minsky human mind contains numerous agencies which compete at any moment in time for imposing their own view/decision about the action to be taken. Depending on instantaneous internal and external contextual factors, these agencies rise and fall in terms of their strength in the competing process. This permanent conflict-like process has always a single winner, or in other words the compromise between two agencies is impossible, hence the name "principle of non-compromise". It is a view that comes in opposition with the usual approaches in cognitive science. In all major cognitive agent architectures (including those presented in Chapter 2), regardless the cognitive theory they follow, decision is modelled through conflict resolution mechanisms. These mechanisms are implemented based on certain profit or utility functions in order to generate a convenient compromise outcome given a set of possible courses of action. Principle of non-compromise was not formally/scientifically validated by AI or cognitive theories. However, the existence of internal agencies within the mind of human individuals and the 'non-compromising' competition between them have been intensely studied in psychological theories under the name of "dialogical self". Dialogical self considers human mind, in a similar manner to Society of Mind, to be a collection of voices trying to be heard and to impose their own way of action. According to Lysacker, dysfunctions observed by psychiatric practitioners in patients with multiple personality, dissociative disorders or decision-making disabilities, are generated by a "collapse of the dialogical self" (Lysaker & Lysaker, 2002). Also, Hermans (Hermans, 2002) sees this as an "organisational problem" of the self. He considers that if voices are unable to "non-compromise" they start to coexist (overlap) in the decision-making, generating ambiguous or improper courses of action.

3.1.2. The k-line theory of memory

The "k-line theory of memory" is also a metaphor, in which the memory of past actions is seen as the "knowledge base" from which component entities of the agencies in the Society (of Mind) are made of. Perhaps there is no better way of explaining this but citing an entire paragraph from Minsky's book:

> "*You want to repair a bicycle. Before you start, smear your hands with red paint. Then every tool you need to use will end up with red marks on it. When you're done, just remember that red means 'good for fixing bicycles'. Next time you fix a bicycle, you can save time by taking out all the red-marked tools in advance.*
> *If you use different colors for different jobs, some tools will end up marked with several colors. That is, each agent can become attached to many different K-lines. Later, when there's a job to do, just activate the proper K-line for that kind of job, and all the tools used in the past for similar jobs will automatically become available.*"

Exactly in the way the bicycle tools story says, decision on choosing a certain course of action is based on recalling and re-composing past facts, actions, or images of the world. These are stocked in the long term memory as the result of encountering various life situations (Minsky, 1985, 1991). K-line selection "re-members" only those bits of information available in the long-term memory which are relevant for the current life context which the individual is exposed to. The core of this idea was used in some cognitive agent architectures under the name of "chunking" (Anderson, 1996; Laird, et al., 1987). However its computational instantiations were limited to implementations built on symbolic production rules. Examples are: the rule-based chunking mechanism used for implementing the long-term knowledge base in SOAR (Laird, 2012), and the long term declarative memory included in the meta-cognitive control layer in ACT-R (Salvucci, 2006). Yet, decision-making in both SOAR and ACT-R relies almost exclusively on recalling and selecting facts from long-term knowledge base. In other words, the chunking mechanism ***is*** the decision-making. In Minksy's approach K-line selection is viewed only as a step within the decision-making process. The "re-membering" mechanism feeds multiple agents and agencies of the mind in order to support their state update and their position in the sub-agent competition process.

3.1.3. Agent "heterarchy"

Principle of non-compromise, together with K-line selection of past events are the main constituents for creating what Minksy calls a "heterarchy" of agents within the internal society of the mind. The idea of heterarchy is proposed as opposite to "hierarchy" which is in Minsky's opinion a simple but rather inaccurate and insufficient way of dividing work into simple tasks in order to solve complex problems. Within the human mind an agent hierarchy, can be organised as a tree with each agent having simple things to do: "look up" for instructions from supervisor and "look down" to exercise control over the subordinates. However, Minsky claims that "*hierarchies do not always work*", and argues that in most complex cognitive tasks internal sub-agents must use each-other's skills. Thus, none of them can be above or below (i.e. a supervisor or a subordinate) in a hypothetical hierarchy but rather they are part of a heterarchy. In this heterarchy not only a tree-like structure exists but also loops and cross-connected rings, depending on the context. He suggests that such heterarchical structures must make use of memory in order to regulate their activity. Also, by lacking a strict hierarchy they need a certain mechanism for establishing which agent is the dominant voice at a certain moment in time. Consequently, instead of a fixed control framework that governs a hierarchy, a dynamic regulatory process is suggested. In this process the K-line memory selection and the principle of non-compromise are dynamically governing the heterarchy.

3.1.4. Dynamics of the "Society of Mind" mind – my own narrative exercise

Minsky treats the human mind at a very high level of detail, but keeps all his claims under a highly narrative tone. There is no reference throughout his work to any possible instantiation of his ideas, be it in a computational or non-computational approach. In order to summarise his view and to make a step towards a possible computational instantiation of the whole theory, dynamics of Minsky's heterarchical approach is presented below in an own narrative manner:

- human mind, or its corresponding computational cognitive agent, is not a singular entity but a collection of primitive entities called "agents" organised as in a society-like edifice in a pseudo-hierarchical structure called "heterarchy". Primitive agents can act alone as agents or they can act within a group of agents which is called an "agency". An agent can be part of more than one agency in the same time, and can be part of different agencies in different situations at different moments in time;
- human mind does not possess intrinsic intelligence. Intelligence comes from exposure to real-life situations, through accomplishment of different tasks by using some of the internal agents and agencies. Agents and agencies used for those tasks become specialised and they are stored in long-term memory;
- when facing a real-life situation 'A' the human mind activates a K-line selection process which tries to bring from memory those agents and agencies which in the past were involved in solving situations of type 'A'. Their partial knowledge about situation 'A' is recomposed in order to generate a course of action which will solve 'A';
- however, various agencies may hold different knowledge about situation 'A' and they propose several possible courses of action, which are most of the time in conflict. For each possible course of action the involved agents or agencies manifest their intention to act with a certain intensity called "strength". In this case the principle of non-compromise is activated and only the strongest agent/agency is chosen. The chosen one will forward its proposed course of action to actuators.
- strength of on agent/agency varies in time depending on the environmental context. Thus, a situation 'A' can be solved differently at different moments in time depending on which agent/agency has the highest strength for solving 'A' in that particular environmental context.

Starting from this short narrative exercise that summarises a potential internal dynamics of a Society of Mind agent, in the next sections a computational framework for building such agents will be described in detail. In the following section the narrative summary presented above will be transferred into an architectural schema in order to establish the architectural framework for a "Society of Mind" cognitive agent.

3.2. The proposed SoM agent – general architectural schema

A problem that appears in the process of transferring the narrative description into an architectural schema is the K-line theory of memory. If principle of non-compromise can be represented as it is following step by step Minsky's narrative description, the K-line theory must be slightly altered. The reason for this alteration is that implementation of K-line theory as it is in a potential application equals to recreating the whole long term knowledge base, i.e. the evolution of an individual from the moment of birth to the present time of the simulation. The K-line theory of memory in Minsky's view is in essence a narrative way of explaining the process of formation of human mind (the "self"). It actually describes the emergence of intelligence and cognitive capabilities throughout the evolution of an individual. An artificial agent with this kind of K-line representation should be subject of a tremendous training in advance, before the simulation starts, in order to build in it the desired knowledge base, i.e. the cognitive capabilities of interest.

Attempts to implement this idea exist in various architectural approaches, such as SOAR or ACT-R, which were discussed in previous chapter and mentioned again in a previous section of this chapter. Arguably, the "chunking" mechanism incorporated in SOAR and ACT-R contains the core of Minsky's approach but respond to the need of building the knowledge base with a symbolic, rule-based initialisation. The agent is endowed with predefined production rules which select action courses from long-term memory (knowledge base) depending on the current situation existing in working memory. In this approach, depicted in Figure 3.1, the decision making process is built around the long-term memory. Actually, the process of selecting the appropriate course of action from the long-term memory (knowledge base) is in itself the decision making mechanism of the agent. In this case, the working memory only feeds the decisional process with current contextual data.

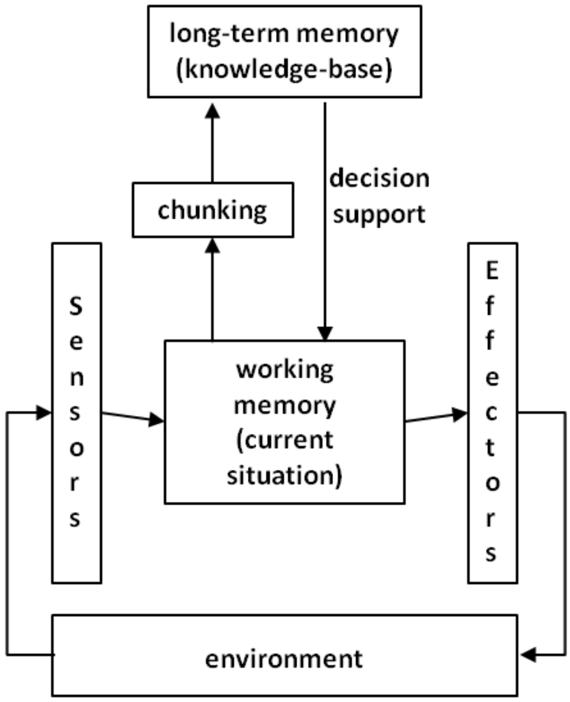


Figure 3.1 A chunking mechanism with separate long-term and short term memory

Arguably, another way of creating the long-term knowledge base is through the agent structure in itself, as in Figure 3.2. The internal structure of an agent and the interactions between its components show what and how the agent processes input information in order to generate output, which in fact accounts for the knowledge base. In this case decision is built around agent's structure: if agent's structure is flexible than the decisional process becomes versatile.

Starting from the above idea, the K-line memory can be understood (implemented) as follows. K-line selection is seen as a set of potential actions that an agency of the SoM agent proposes at current step for the next step – given its internal state and the current state of the environment. Hence, the K-line selector does not select from a set of past actions and tools in order generate an action for the present time, as in the chunking mechanism. It rather selects from a set of "*ready to do*" actions of a sub-agency only that action which is applicable to the current environmental state. As an example, a driver agent in an affective state of advanced anger is assumed. Anger is known to generate actions like speeding, tailgating or cutting. If there is no other vehicle on the street in that specific moment of time, then cutting and tailgating are not applicable to the context. The only valid action to be proposed for actuation is speeding.

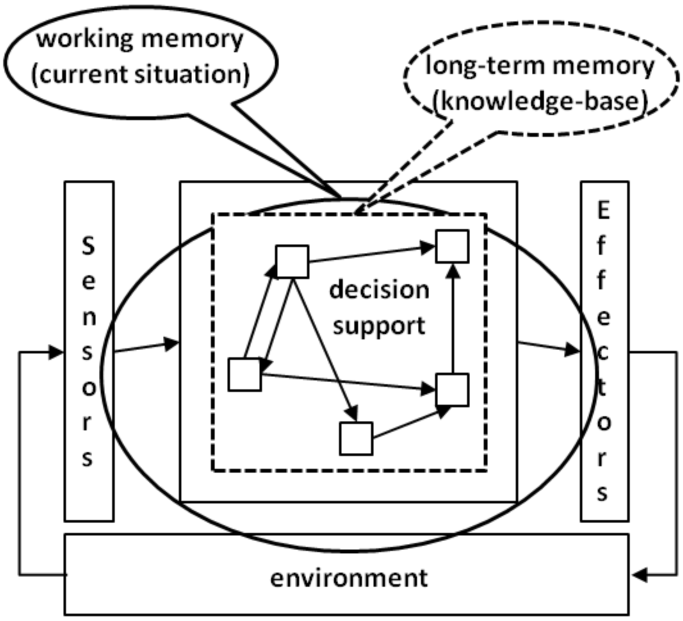


Figure 3.2 Implicit knowledge base through agent's internal structure

Figure 3.3 shows the proposed architectural framework of a potential Society of Mind agent and describes its potential internal dynamics. It includes the modified form of K-line selection, the non-compromise principle and the heterarchy of agents.

The SoM agent consists of a number '*n*' of sub-agencies *A* (A_i; $i=1,n$), each of them containing a set of internal entities (basic agents) *e*. First, an agency A_i senses the environment through its internal entities, by gathering from the larger input dataset *x* (x_i; $i=1,m$) only that information which is relevant to it (A_i). Then, its entities interact and update their states and through that the state of the entire agency. As a result of interaction of its internal entities (basic agents) the agency A_i is able to propose a set $K\text{-}line_i$ of actions which are usually the result of that particular interaction of internal entities *e*. However, not all these potential actions are relevant to the current context: the state of the environment and the state of the overall SoM agent within the environment. Thus, a K-line selector chooses from the $K\text{-}line_i$ set of potential actions only that action (or subset of actions) act_{Ai} which is strictly related to the current context. This action will participate in the non-compromising competition with the actions proposed by other sub-agencies. Once all agencies finish their "preparation" and the K-line selector selects the candidate action for each agency, these actions participate in a bid. According to a certain bidding rule/strategy only one course of action will win and will be transferred to effectors to actuate the action in the environment.

From a Perception-Action perspective the SoM agent architecture is still in a classic SDA loop (discussed in Chapter 2). The SoM agent *senses* the environment through its agencies, *decides* about a course of action and *acts* upon the environment through effectors. However, the internal design of the decisional stage suggests an approach such as SPNA (Sense Prepare Non-compromise Act). This approach is preffered based on the fact it shows in a more appropriate way the real dynamics within the Perception-Action cycle. Indeed, all the internal processing which happens inside each agency before the competition stage reflects the way from the cognitive appraisal (sensing) to a state of awareness (preparedness). In this state the overall SoM agent becomes aware and prepared for multiple courses of action and their effect. For this reason "prepare" stage can be considered as a first step in the decisional process. Once the agent is aware of the multiple courses of action the next stage of the decisional process starts: the competition. Following Minsky's principle of non-compromise this stage can be named the "Non-compromise" step.

From the cognitive capabilities architectural design point of view the SoM architecture can be treated as a non-hierarchical hybrid architecture in which numerous competing cognitive processes can be simultaneously considered. Arguably, the resultant SoM agent is in the first place much more versatile than any other hierarchical architecture due to its very design.

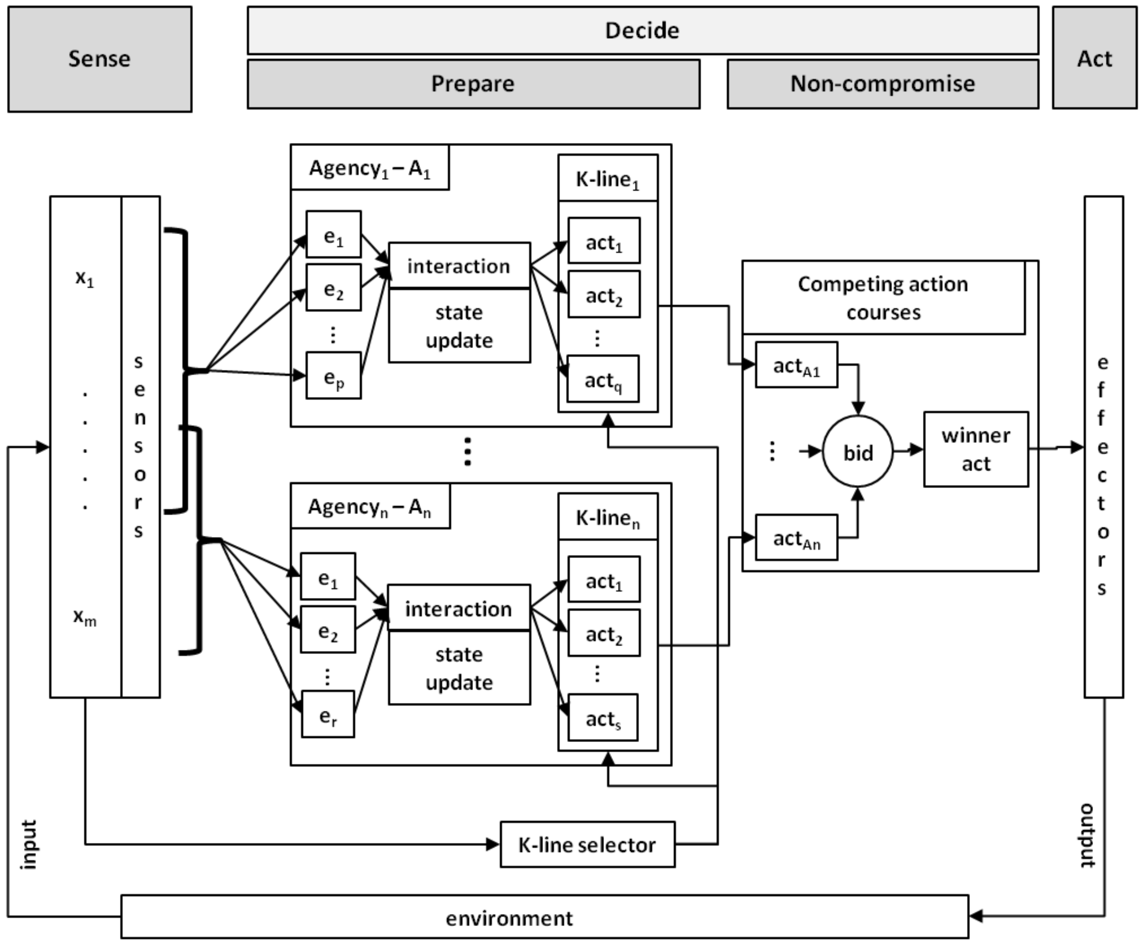


Figure 3.3 A generic "Society of Mind" architectural schema

First, the SoM agent should be able to cope with complex problems which combine all *n* types of cognitive processes corresponding to the *n* agencies (A_1 to A_n). Hence, the range of cognitive capabilities supported by an SoM agent is given by the number and types of agencies implemented in a potential instantiation. Second, a complex SoM agent consisting of *n* agencies can reduce its capabilities when needed by inhibiting one or more of its agencies when the current tasks do not involve certain cognitive processes. From an implementation point of view this could be done either by suppressing the whole cognitive processing branch for the unused capability (e.g. suppress one agency from the SPNA loop). Even simpler, this can be done by obstructing the participation in the bidding process through enforcement of a disqualifying value. It can be then assumed that an agent of this type has high portability over various fields of activity and could solve, at least in a certain extent, the problem-dependence drawback of existing hybrid architectures. In the same time, it also offers a certain theoretical support, despite being based on a metaphorical understanding of human mind, rather than on a well-established cognitive theory.

3.3. A three tier RRA SoM architecture – detailed view

Previous section presented a general architectural schema that transfers Minsky's narrative approach on human mind and cognition into a potential architectural instantiation. The general architectural schema emphasises on the ability to insert in an SoM agent as many cognitive functions are needed for a specific task, in the form of agencies, in order to obtain the desired behaviour. However, in this thesis the agencies of interest are those representing the fundamental categories of cognitive capabilities discussed in detail in the literature review chapter: reactive, rational and affective. Arguably, an agent design containing these three agencies would cover a very wide range of cognitive processes, allowing high versatility and high representation power of human behaviour in virtually any task required by human activity.

Hence, this section proposes a three tier Reactive-Rational-Affective (RRA) SoM architectural approach. The approach starts from the assumption that a human being can be in most of the situations encountered in day-to-day activity in the position of choosing between the following three major types of decisions:

- **a reactive decision:** an agent (human) MUST react to sudden, unexpected changes in the environment/context by taking rapid avoiding actions in order to escape from an immediate danger;
- **a rational decision:** an agent (human) takes rational decisions in order to fulfil goals related: to own (or others, or environmental) safety, to abiding social/legal norms and regulations, or to following a certain prescribed and generally accepted profit or utility function;
- **an affective decision:** an agent (human) takes affective decisions based on the current emotional state, regardless the social norms and

regulations, the safety, the annoyance produced to others or the unwanted effects on the environment.

Figure 3.4 shows the three tier RRA SoM non-hierarchical hybrid architecture which is based on the general SoM architectural schema described and depicted (Figure 3.3) in Section 3.2. Internal dynamics of the RRA SoM agent is similar to the one explained for the general case, and for this reason only a generic representation of the three agencies is shown in Figure 3.4. Given that the RRA SoM architecture is a particular case of the more general SoM schema, a subsequent RRA SoM agent should keep the features assumed for SoM agents: versatility, power of representation of a wide range of human behaviours and high portability. Regarding the portability an RRA SoM agent should also be able to suppress one or more of its agencies, thus becoming a simpler agent, e.g. a reactive-deliberative hybrid, or an affective-only agent, or a reactive-only agent.

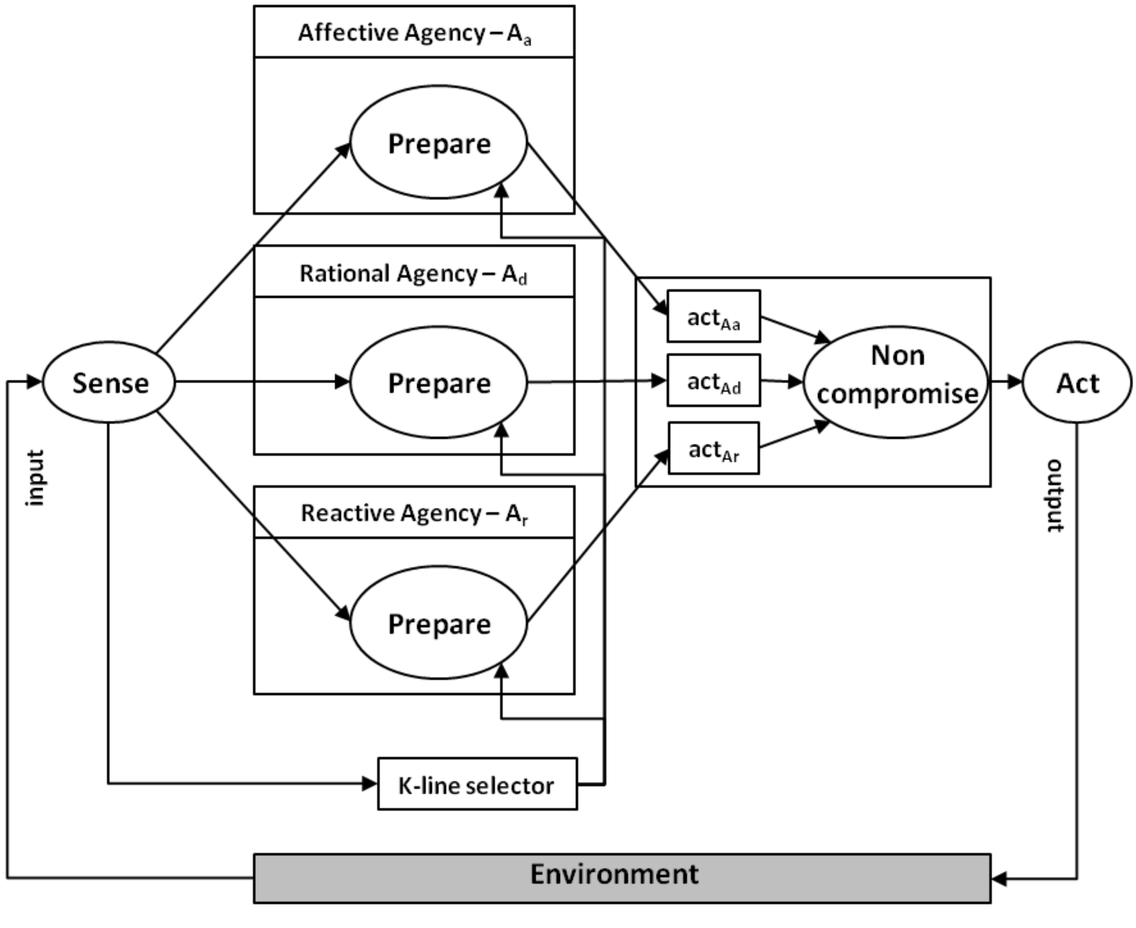


Figure 3.4. A three tier RRA "SoM" agent architecture

In addition to RRA assumption made above, another addition must be made to Minsky's SoM metaphor, in order to obtain a more versatile agent. As discussed in Chapter 2, there is another aspect of human mind and cognition that was not addressed by any of the existing complex cognitive architectures: the innate individual biases and behavioural propensities. Sun and colleagues mention that CLARION architecture could support implementation of such aspects, due to the existence of a meta-cognitive control layer (Sun, 2006; Sun, et al., 1999). However, they do not give any detail about what specific cognitive features can create such biases and propensities. Also, they neither signal any existing implementation of such features nor mention any intention to do so in the future.

Figure 3.5 shows how non-contextual mental factors, either innate or acquired, can be added to usual contextual data acquired from sensors in order to update the current state of an agent. Integration of this concept in the RRA SoM architectural approach is presented in Figure 3.6.

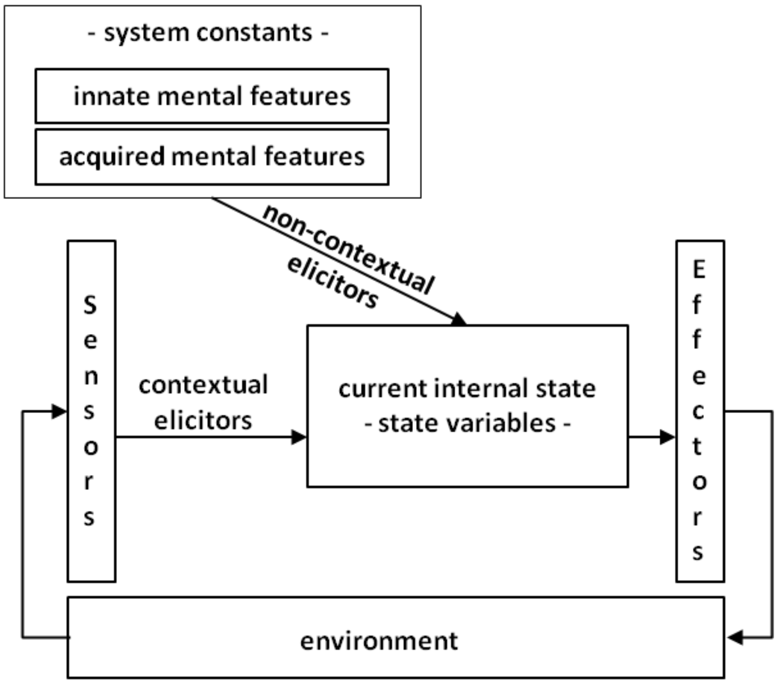


Figure 3.5 Contribution of non-contextual elicitors to decision-making

Starting from this view on RRA SoM agent, as a possible instantiation of Minsky's "Society of Mind" theory, the following paragraphs will present in detail each element of the proposed architectural schema. Also potential methodologies for implementation of these elements will be suggested.

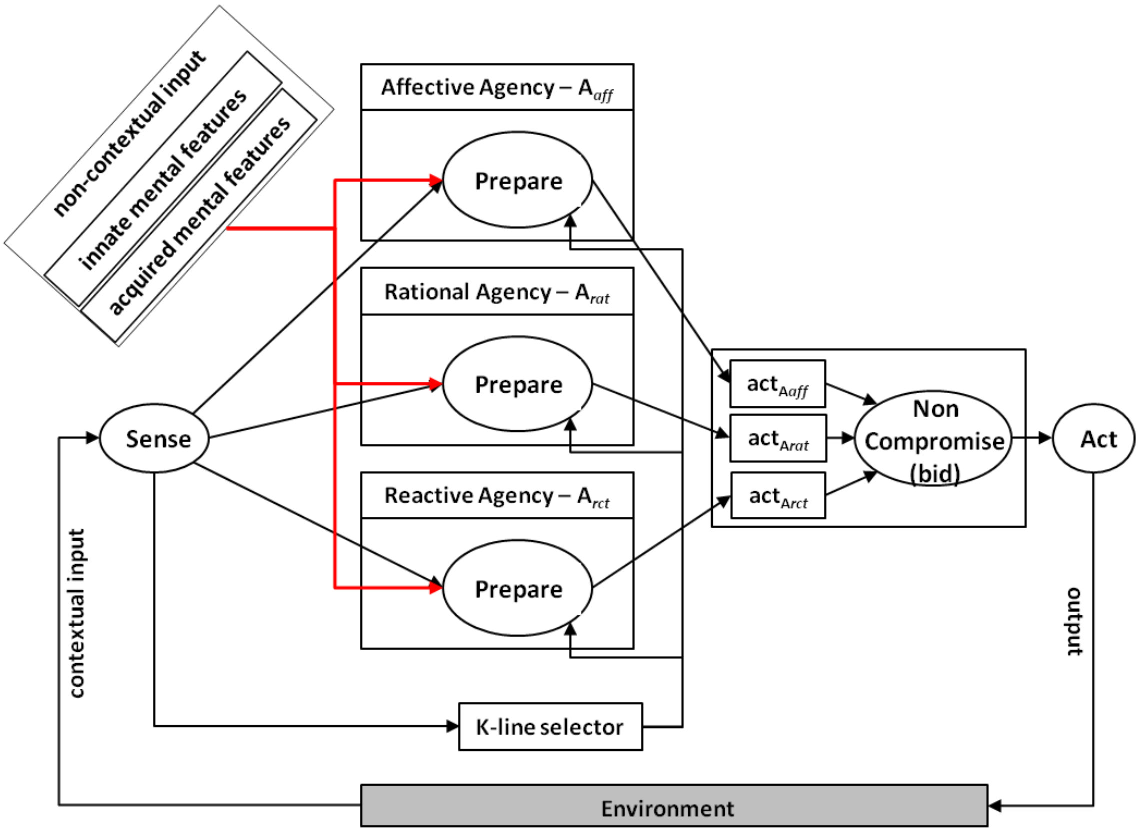


Figure 3.6 Complete RRA SoM architectural scheme with inclusion of innate biases and behavioural propensities.

3.3.1. The innate and acquired behavioural propensities.

In most real-life situations, apart from decision based on contextual inputs coming from either the environment or the interaction with other individuals, humans also display constant tendencies to act in certain ways. These tendencies may yield from innate elements due to genetic inheritance or can be acquired from education or social environment and lifestyle. They influence the way the individual reacts to environmental stimuli and subsequently, they influence the decision-making and further, the resultant behaviour.

3.3.1.1. *Innate propensities – personality and individual differences*

Innate propensities (or biases) were intensely studied by the personality theory, under the wider umbrella of theory of "individual differences". However, computational instantiations of the numerous findings in this research field are very few in the modelling of human activity systems. In the field of complex cognitive agents, they merely do not exist, as discussed in the literature review chapter.

Personality denotes that unique mix of characteristics or qualities which forms the distinctive character of an individual and is considered as arising from within the individual as an innate matter. The in-depth discussion provided in Chapter 2 showed that dimensional approach on personality, i.e. the trait theory, stood out from all the approaches proposed through a very long history of personality research. Among the dimensional approaches the BIG 5 model (Goldberg, 1990) is the most prolific and universally recognised, with proved applicability and predictive capabilities over multiple fields of activity. This fact recommends it for usage with the RRA SoM architecture and the next paragraphs will emphasise on its use. However, other dimensional approaches, such as the three factor model (Eysenck, 1991), or even non-dimensional approaches can be chosen if they are useful for a particular RRA SoM agent implementation. For the scope of this thesis the description will be limited to dimensional approach using as an example (and recommendation) the BIG 5 model.

Dimensional models assume the existence of a number '*n*' (*n*=5 five for BIG 5 model) of fundamental bipolar and orthogonal dimensions of the personality called traits or factors. Any human being is situated in a continuous range between low and high values for each of the factors, as in an '*n*' dimensional personality space.

Figure 3.7 depicts in a graphical manner the dimensional approach on personality. Mathematically this narrative description can be expressed as an *n*-dimensional tuple $P(P_1, P_2, \dots, P_{\boldsymbol{n}})$ with $P_i \epsilon (-\infty; \infty)$, where P_i are the personality factors. From a computational point of view various implementations can be suggested depending on the specific application. One approach could be the usage of a convenient limitation of infinite values of personality factors P_i to finite real intervals $P_i \epsilon [-P_{max}; P_{max}]$ or integer sets $P_i \subset \{x | -P_{max} \leq x \leq P_{max}, x \in \mathbb{Z}\}$. In a different approach, the unlimited intervals can be scaled to a real interval (0;1) by using a sigmoid function. The sigmoid can be used either in the canonical form $P_i = S(P_i) = \frac{1}{1+e^{-P_i}}$ or in the more general setup as shown in (3.1), depending on the implementation needs.

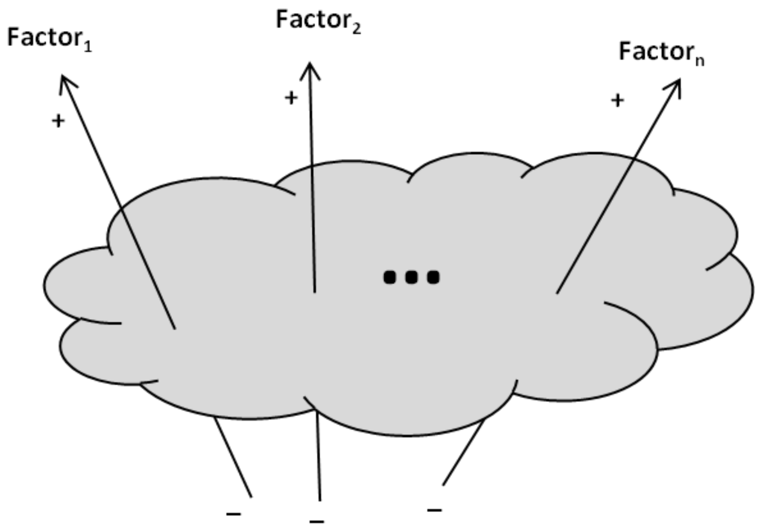


Figure 3.7 Personality – dimensional approach

$$\boldsymbol{S(x) = \frac{1}{1 + e^{-a(x-b)}}; a, b} \in \mathbb{R} \tag{3.1}$$

The values of personality traits/factors are considered system constants and are assigned to the agent at the beginning of simulation. These values are fed to internal agencies during the simulation together with variable contextual information coming from sensors, and participate in the state update process implemented in each agency.

For an intuitive example of how innate unchangeable personality traits affect decision making in real-life situations, the BIG 5 model will be considered with its five fundamental factors: openness, conscientiousness, extraversion, agreeableness and neuroticism. Based on the discussion on personality from Chapter 2, it can be assumed that individuals which score high on neuroticism and low on agreeableness will tend to build up aggresivity related emotions such as anger. As a result, decision-making will be biased towards affective decisions situated on the negative behavioural side (aggresivity, disrespect, unsafe behaviour, risk appetite). If on the contrary an individual has high agreeableness and low neuroticism, it can be assumed that such a personality type would bias the decision-making towards an increase in the strength of rational voice. Thus, a rational decisional outcome situated on the positive behavioural side is expected (rule abiding, politeness, safety oriented etc.). However, the influence of personality traits on decision-making can vary widely with the type of application the RRA agent is built for. The present example is only used as an attempt to describe in an intuitive manner the process standing at the basis of innate behavioural propensities.

3.3.1.2. Acquired propensities. Moods, social norms, subcortical pathways

Acquired behavioural propensities represent a very wide problem, treated by numerous fields of activity and promoting countless theoretical and practical approaches. A detailed investigation of these aspects is outside the scope of this thesis. It was included in the general architectural design for completeness purposes, since such aspects cannot be left out of discussion. Implementation of acquired behavioural propensities is indeed very closely related to the particular purpose of an RRA SoM agent. This makes difficult even to summarise a set of available methodological options. However, a brief discussion will be provided in the following paragraphs.

Acquired propensities are not constant throughout an individual lifetime and they do not arise from its genetic inheritance, therefore they are not similar to personality. Still, with regard to the timeframe of a task solved by a potential RRA SoM agent they can be considered system constants, unless the focus of agent's implementation is exactly on the study of such aspects. Though, the latter case is not treated here, and the discussion will continue for the situation in which acquired biases can be treated as system constants.

The most common elements which can create acquired behavioural propensities and decisional biases were investigated as part of numerous socio-psychological studies thoroughly reviewed in VanBreda's book on individual resilience (VanBreda, 2001). It is largely believed that they are coming from several directions such as individual's social proximity (e.g. family, social network), education, religion etc. They are the so-called life-style related propensities. However, apart from the sociological view there is also a different approach, more related to physiology of the individual, described in neuro-psychology literature as the "sub-cortical pathway" (Critchley, 2003). According to this, cases are when human decisions cannot be explained in either of the rational or emotional approaches. It is believed that chemical/neural reactions act at physical levels of the human brain to produce mood-like mental states, which influence decision regardless the internal or external, contextual or non-contextual, and innate or non-innate aspects. From a cognitive-affective point of view moods are considered as part of the emotional system (Lisetti, 2002), being described as emotions with a longer temporal range. However, in the neurological approach their emergence is different from that of emotions. From an agent perspective though, in the light of RRA SoM architecture discussed in this chapter moods can be considered, together with the socially generated biases, as acquired propensities.

As a conclusion it is important to note that the proposed RRA SoM architectural approach allows implementation of acquired propensities, however a specific implementation methodology or a set of potential methodologies are not treated at the moment.

3.3.2. The affective agency.

From a structural point of view affective agency is based on the main constituents of cognitive-affective system: emotions. Emotional system presents certain similarities to personality, through that it consists of a set of emotional traits, also called features or simpler – '*emotions*'. However, emotions are not innate and not at all constant as traits are in the personality system. On the contrary they are highly variable in time, accounting for the real-time context generated affective decision-making. In the literature, emotions were studied from two perspectives: dimensional and discrete (non-dimensional), both of them being discussed in detail in Chapter 2.

From the point of view of the proposed RRA SoM architecture this thesis will focus on the dimensional approach for two reasons. First, it is highly compatible with dimensional approach on personality and second, dimensional approach seems to have been embraced by a larger number of scholars and practitioners in the last years. However, the implementation methodology presented in following paragraphs can be still used in relation to discrete emotions if the specific tasks to be solved by RRA SoM agents require to do so.

Dimensional approach considers there are several fundamental emotional directions, bipolar and continuous that can represent the variety of human emotional responses. There is almost a consensus that the emotional space consists of three dimensions, as shown in Figure 3.8: valence, activation and dominance (Bolls, et al., 2001; Grimm, Kroschel, Mower, et al., 2007; Lang, et al., 1995). However, in order to keep a general approach an emotional space with *m* dimensions will be used.

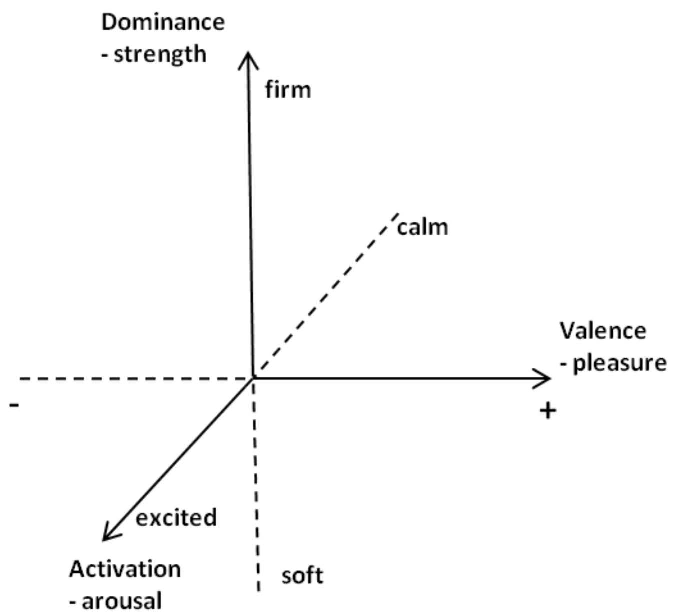


Figure 3.8 Emotions – dimensional approach

Similar to dimensional approach on personality individuals can be for each of the emotional dimensions somewhere in the continuum between the two bipolar extremes. Position in the emotional space is an instantaneous one which varies in real-time according to the variation of affective states, unlike the position in personality space which is constant over time.

Mathematically a similar approach to that of personality can be followed, i.e. emotional space can be represented as an m-dimensional couple $E(E_1, E_2, \ldots, E_{\boldsymbol{m}})$ with $E_i \epsilon (-\infty; \infty)$, where E_i are the primitive emotions. Computationally, the problem generated by emotional infinite intervals can be also solved following the same approach used for personality. First, use a convenient limitation of infinite values of emotions E_i to finite real intervals $E_i \epsilon [-E_{max}; E_{max}]$ or integer sets $E_i \subset \{x| - E_{max} \leq x \leq E_{max}, x \in \mathbb{Z}\}$. Second, the unlimited intervals can be scaled to a real interval (0;1) by using a sigmoid function either in the canonical form $E_i = S(E_i) = \frac{1}{1+e^{-E_i}}$ or in the more general setup shown in (3.1).

Figure 3.9 shows the proposed the internal dynamics of affective agency. The long term unchangeable innate personality traits and the short term context-generated emotional states interact. This interaction updates the affective state and generates the consequent strength for the proposed affective action courses.

Emotions E_i are initialised with values which are relevant for the specific task the RRA SoM agent is implemented for, and they are updated as in the intensity-decay approach described in Velasquez's Cathexis model (Velásquez, 1997). In his model Velasquez assumes that intensity of an emotion varies according to a predefined function in every update cycle (simulation time-step). If there is no inhibitory or excitatory input, the emotion decay and after a few cycles it becomes inactive, as in equation (3.2). Also, it should be noted that equation (3.2) only presents the general concept of an intensity-decay function for emotion update, as in Cathexis model proposed by Velasques. The intensity-decay function is presented in order to explain the concept of intensity-decay. For this reason, no particular values are given to various paramatres of the function.

$$\boldsymbol{E_{it}} = \boldsymbol{f_{CAT}} \left\{ \boldsymbol{D}(\boldsymbol{E_{it-1}}) + \sum_{\boldsymbol{k}} \boldsymbol{L_{ki}} + \sum_{\boldsymbol{l}} \boldsymbol{G_{li}} \, \boldsymbol{E_{lt}} - \sum_{\boldsymbol{m}} \boldsymbol{H_{mi}} \boldsymbol{E_{mt}} \right\} \tag{3.2}$$

where E_{it} is the intensity of emotion i at time t, $D()$ is the decay function of emotion i, L_{ki} is the elicitor k of emotion i, G_{li} is the excitatory gain that emotion l applies on emotion i, H_{mi} is the inhibitory gain that emotion m applies to emotion i and f is the function that places the emotion i between 0 and its maximum value.

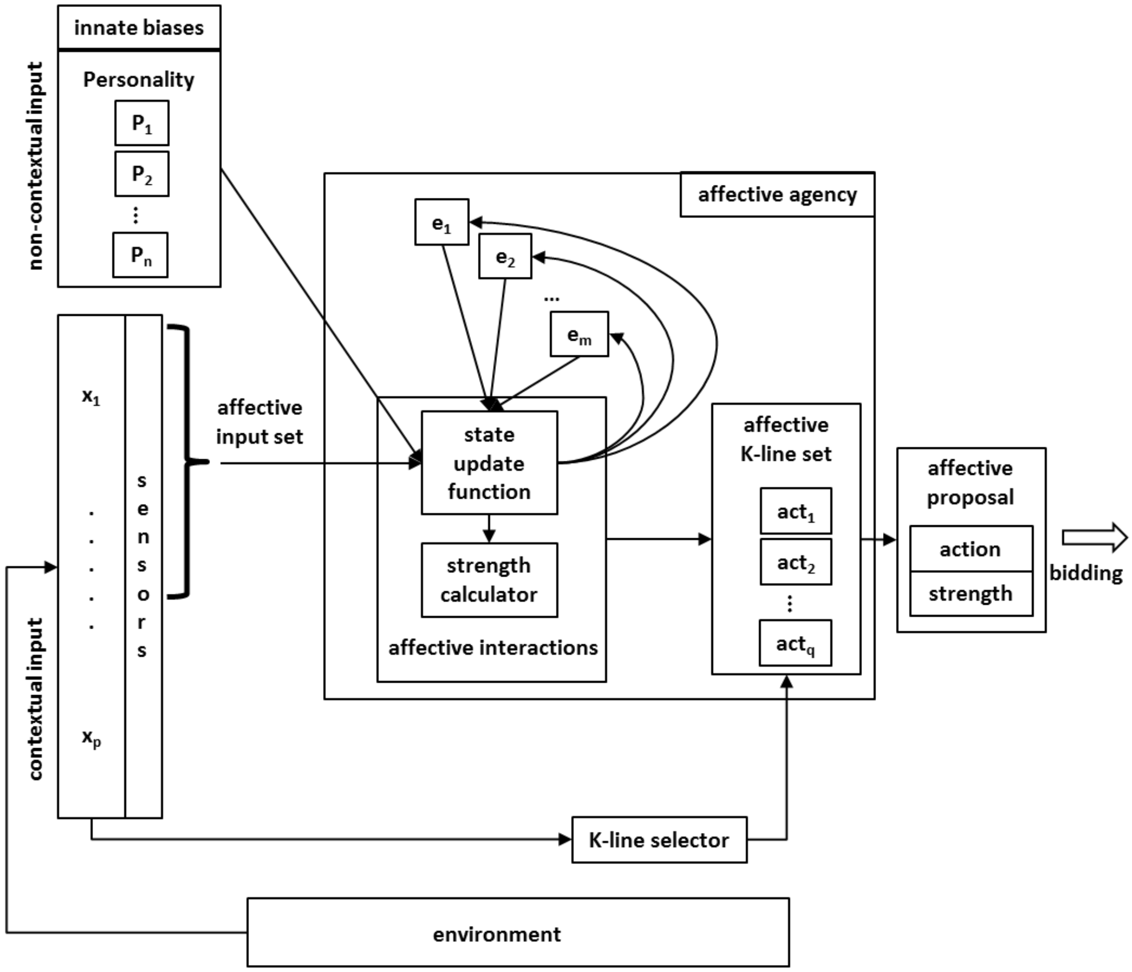


Figure 3.9 The affective agency – internal dynamics

However, Velasquez's model only treats the emotional update process, whereas update of the affective agency also involves participation of personality traits vector P and affective input vector X. The resultant state update function of affective agency, as well as the sequences involved in the decisional process, is described below.

Sense:

At simulation time-step t_n the affective agency prepares for potential actions based on the affective set of contextual inputs $X_{aff}(X_1, X_2, \dots, X_{paff})$ and also taking into account the unchangeable values of personality traits and values of emotions at t_{n-1}.

Prepare:

The ‘prepare’ stage consists of three steps. First, the affective agency updates the values of emotions using the following input data: contextual data from environment, selected personality traits with a contribution to affective decision, and previous values of emotions. Second, based on the newly calculated values of emotions, the value of action *strength* is computed. Third, the K-line selector selects from the affective action set that action which is applicable to the current state of environment. This action, together with the *strength* value will continue to the non-compromise stage.

Step 1 – update of emotional state. Starting from CATHEXIS model proposed by Velasquez, the emotional update function can be seen in general as in as in (3.3).

$$\boldsymbol{E}(\boldsymbol{t_n}) = \boldsymbol{f_{update}}\{\boldsymbol{P_{aff}}, \boldsymbol{f_{CAT}}(\boldsymbol{t_n}), \boldsymbol{X_{aff}}(\boldsymbol{t_n})\} \tag{3.3}$$

where E is the vector of emotions, P_{aff} is a subset of the personality trait vector containing those traits involved in affective processes, f_{CAT} is an appropriate intensity-decay function for emotions update which includes the short memory effect (i.e. emotional states at time t_{n-1}) and X_{aff} is a subset of the contextual input vector X consisting of those inputs applicable to affective agency.

Step 2 – update of strength. After the emotional states are updated, they are used together with personality traits and contextual inputs for transferring the updated emotional state into a *strength* level. The strength value will further participate in the bidding process included in the *non-compromise* stage of decision. A general update function of the *strength* is presented in (3.4).

$$\boldsymbol{aff_{strength}(t_n) = f_{strength}\{P_{aff}, E(t_n), X_{aff}(t_n)\}} \tag{3.4}$$

where P_{aff} is a subset of the personality trait vector containing those traits involved in affective processes, E is the updated vector of emotions, and X_{aff} is a subset of the contextual input vector X consisting of those inputs applicable to affective agency.

Step 3 – K-line selection. According to the proposed internal structure of affective agency which accounts for long-term knowledge, and to the update functions which account for short-term memory, a set of actions is proposed for the current step. These actions are not actuated immediately, as they are. They only create the premises/preparedness for generating in the physical domain those effects which are related to them. The K-line selector chooses from the existing actions only that (or those) one which can be applied to current situation. Selected action is then proposed together with its strength as a competitor in the bidding process of the non-compromise stage. A detailed discussion about potential K-line selector implementation will be provided later in this section in a dedicated subsection.

3.3.3. The rational agency

In the most common view, described by Russel and Norvig (Russell & Norvig, 1995), rational agents select for each possible perception sequence an action that is expected to maximise a certain utility function.

The major architectural approaches and implementation methodologies for rational agents are built around the paradigm of deliberative architectures discussed in detail in Chapter 2. They have their roots in various rational choice theories, starting from the classical AI assumption that agents such as people or organisations act rational. However, what rationality is stays nevertheless in the eyes of the beholder, namely the agent designer who decides on what real-life substrate the respective instantiation of rational choice theory is built on. Hence, what is chosen by the agent as an optimal way of action from a rationality point of view could be the result of embracing social norms and accustomed habits. It could also be the result of obeying laws and regulations or even invariant prescribed actions that fall, in the view of system designers, into an assumed concept of rationality.

From the RRA SoM architecture point of view, any of the existing implementation methodologies can be used. Implementations with condition-action rules as in the major symbolic AI approaches, such as reflex, goal-based or utility-based models, are without any doubt usable and useful for a wide range of problems. However, in order to keep the discussion for the proposed SoM approach as general as possible, non-symbolic implementation approaches must be also taken into account.

In the most general view an approach similar to the one proposed for affective agency could, arguably, cover a wide range of implementation methodologies. As discussed in Section 3.3, the assumed rational behaviour of an RRA SoM agent should cover more than just the fulfilment of a profit/utility function. It should also cover prescribed actions which aim towards agent's own safety, or towards abiding social or legal norms. Figure 3.10 shows the potential internal dynamics of such a rational agency. The rational behaviour is grouped into a generic set of rational elements called *rationals* which create a virtual space of rationality, in the same manner personality traits and emotions are seen in personality and emotional spaces, respectively.

The set of potential rational elements contained in rational agency's structure is denoted as $R(R_1, R_2, \dots, R_{\boldsymbol{m}})$ where $R_i \epsilon (-\infty; \infty)$ and R_i are the primitive *rationals*. Similar to personality and emotions, the computational problems related to infinite intervals can be also solved by using either convenient limitation of infinite intervals to finite real intervals $R_i \epsilon [-R_{max}; R_{max}]$ or integer sets $R_i \subset \{x| -R_{max} \leq x \leq R_{max}, x \in \mathbb{Z}\}$, or sigmoid scaling to real interval (0;1) using canonical sigmoid $R_i = S(R_i) = \frac{1}{1+e^{-R_i}}$ or the general sigmoid expression shown in (3.1).

Even if an approach like this does not exist in the literature as a formal representation of rational processes, it is very useful for the RRA SoM architecture for being able to represent both short-term memory and long-term knowledge base. Arguably, any of the existing implementation approaches can fit into this generic description. It can be a set of explicit condition-action rules, an equation-based approach, or even simpler a prescribed/fixed decision-making process assumed to be rational.

Starting from the above discussion, the two main stages of the rational agency internal dynamics, *Sense* and *Prepare*, are described in the following paragraphs.

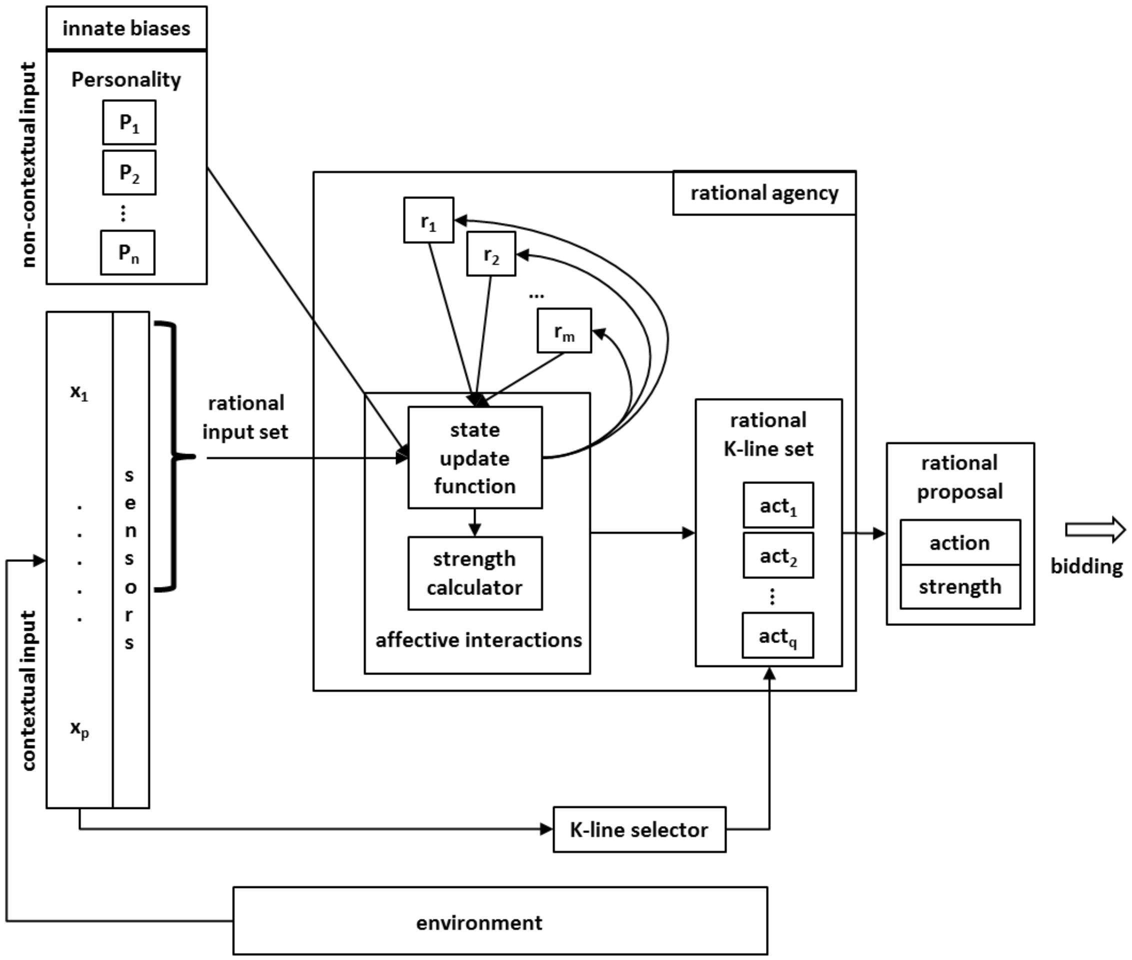


Figure 3.10 The rational agency – internal dynamics

Sense:

At simulation time-step t_n the reactive agency prepares for potential actions based on the rational set of contextual inputs $X_{rat}(X_1, X_2, \ldots, X_{prat})$ and also taking into account the unchangeable values of personality traits and values of emotions at t_{n-1}.

Prepare:

The 'prepare' stage consists of three steps. First, the rational agency updates the values of *rationals* using the following input data: the contextual data from environment, the selected personality traits contributing to rational decision, and the previous values of *rationals*. Second, based on the newly calculated values of *rationals*, the value of action *strength* is computed. Third, the K-line selector selects from the rational action set that action which is applicable to the current state of environment. This action, together with the *strength* value will continue to the non-compromise stage.

<u>Step 1</u> – update of rational state. A generic rational state update function is shown in equation (3.5):

$$\boldsymbol{R(t_n) = f_{update}\{P_{rat}, R(t_{n-1}), X_{RAT}(t_n)\}} \tag{3.5}$$

where $R(t_n)$ is the vector of *rationals* at current time step, P_{rat} is a subset of personality traits vector containing those traits involved in rational processes, $R(t_{n-1})$ is the vector of *rationals* at previous time step accounting for short memory effect and X_{rat} is a subset of the contextual input vector X consisting of those inputs applicable to rational agency.

<u>Step 2</u> – update of strength. After the *rationals* are updated, they are used together with personality traits and contextual inputs for transferring the updated rational state into a *strength* level. The strength value will further participate in the bidding process included in *non-compromise* stage of decision. A general update function of the *strength* is presented below in (3.6).

$$\boldsymbol{rat_{strength}(t_n) = f_{strength}\{P_{rat}, R(t_n), X_{rat}(t_n)\}} \tag{3.6}$$

where P_{rat} is the subset of personality traits vector containing those traits involved in rational processes, R is the updated vector of *rationals*, and X_{rat} is the rational subset of contextual input vector X consisting of the inputs applicable to rational agency.

Step 3 – K-line selection. According to internal structure of affective agency which accounts for long-term knowledge, and to update functions which account for short-term memory, a set of actions is proposed for the current step. The K-line selector chooses the action set only the action(s) applicable to current situation. The selected action is then proposed together with its strength as a competitor in the bidding process of the non-compromise stage. A detailed discussion about potential K-line selector implementation will be provided later in this section in a dedicated subsection.

3.3.4. The reactive agency

AAs explained earlier in this section reactive agency models agent's reaction (reflex) to extreme danger. An agent must react to sudden, unexpected changes in the environment/context by rapidly taking avoiding actions in order to escape from an immediate danger.

The most common implementation methodologies for reactive agents are concentrated around rule-based approaches discussed in Chapter 2: standard condition-action reflex agents, subsumption architectures or agent network architectures.

The condition-action approach is the preferred methodology for RRA SoM agent, in order to describe the action choice. This approach is preferred for its consistency with the strength update, which can be modelled in the condition-action paradigm as well, as explained below.

Figure 3.11 shows the internal dynamics diagram of reactive agency and presents its main constituents, as well as their interaction.

Sense:

At simulation time-step t_n the reactive agency prepares for potential actions based on the reactive set of contextual inputs $X_{rct}(X_1, X_2, \ldots, X_{prct})$. Unlike the other two agencies, reactive agency is not influenced by personality, and also does not contain any short-term memory effect. Thus, its states at previous time steps do not participate in the decision-making process.

Prepare:

The agency is implemented in a rule-based approach in which the agency compares the current context with a finite set of situations stored in its long-term memory. These situations belong to the category of 'danger' and impose immediate avoiding/protective actions.

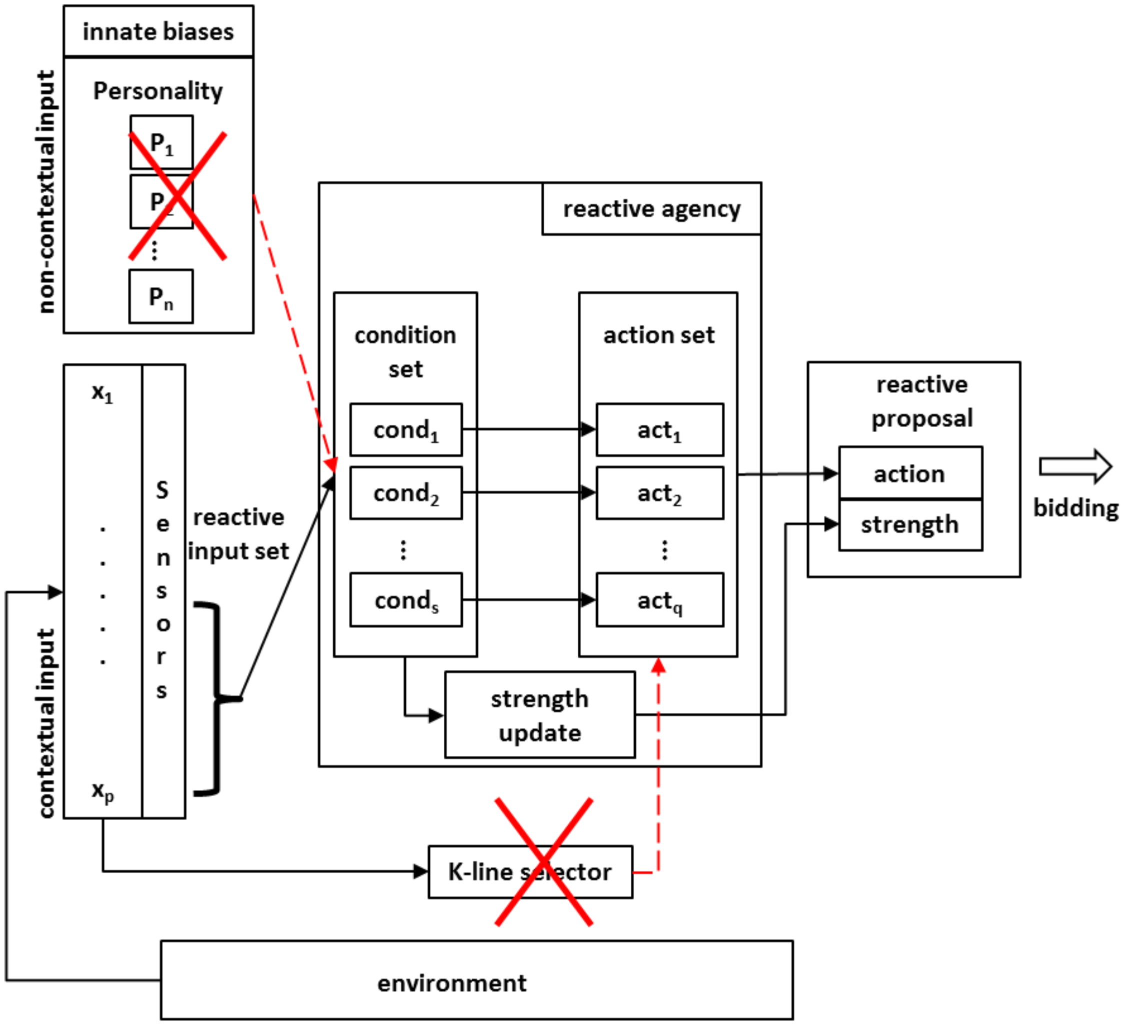


Figure 3.11 The reactive agency – internal dynamics

If the current context is identified as one of these situations then a corresponding predefined action course is proposed for participation in the non-compromise stage. This action should dismiss any other competing action course, be it rational or emotional, since it involves, potentially, survival of the individual. This is reflected in the value of *strength* associated with the proposed reactive action, which must be set to a level that produces 100% chances of winning in the bidding process.

If on the contrary, the current context is not identified as being in 'danger' category, the situtation can be modelled by enforcing a disqualifying value of the *strength*, which forces the reactive agency to lose the bidding.

Also, due to the condition-action approach, K-Line selection is also not applicable for reactive agency since all possible courses of action are properly determined in the set of condition-action rules.

3.3.5. The K-Line selection.

The most appropriate K-Line implementation approach for the proposed RRA SoM is perhaps the one used to implement the chunking mechanism in SOAR and ACT-R. The context is nevertheless different, as explain earlier in this chapter. Chunking brings from long-term memory to working memory chunks of knowledge applicable to current context. As opposite, the K-Line mechanism attempts to select from the knowledge generated by current state of an agency those action courses which fit the context. Despite representing different points of view, the mechanisms are highly similar in their dynamics, hence they could easily share the implementation methodology.

Thus, the K-Line selection process can be modelled in a condition-action rule-based approach. For each agency a set of conditions can be embedded in agent's knowledge base and called after agency's internal state is updated and the set of possible actions is generated. The set of possible actions is connected to the set of predefined conditions so that an action suitable for current context will be selected to participate in the bidding process.

3.3.6. The "non-compromise" stage – bidding.

The bidding process, as part of the non-compromise decisional stage, is a key component of the proposed RRA SoM agent. It represents the instantiation in an architectural framework of the principle of non-compromise, metaphorically described in Minsky's Society of Mind. The proposed RRA SoM architecture presents three agencies which compete for imposing their decision on next action to be done. Instead of searching for a utility based compromise, like in other approaches, the agencies participate in a competition with only one winner. Thus, a bidding process (i.e. an auction-like process) must be implemented instead of a compromise-based conflict resolution mechanism.

A question immediately arising from this fact is what kind of bidding strategy is appropriate for usage within the proposed RRA SoM architectural framework. Existing literature in the field describes two major approaches on bidding/auction strategies: truthful and non-truthful (Cary, et al., 2007).

In truthful strategies participants bid with their true valuation of the goods they are competing for. Truthful bidding strategies are built around Vickrey's strategy (Maes, 1994; Vickrey, 1961) and its generalisation known as Vickrey-Clarke-Groves (VCG) mechanism (E. H. Clarke, 1971; Groves, 1973; Vickrey, 1961). Both are based on the classical English auction used since the 17th century, in which the participant with the highest bid wins and purchases the goods at a price equal to his/her bid (first-price auction). Vickrey and VCG are different from English auctions in terms of auction procedures, and they are known as second-price auctions. They assume that the highest bidder wins but pays the price of the second highest bidder (highest non-winning bid for VCG generalisation). However, the strategy is similar and regardless the price paid, the goods go to the highest bidder. A major limitation of these approaches is the fact they mainly consider auctions in which participants only compete for a single indivisible good.

In non-truthful strategies participants do not bid with the true value of the goods they compete for, in an attempt to maximise various profit functions or win in the long run as opposite to winning the current bid. Thus, they must use more complex strategies which involve extra-capabilities, including knowledge of bidding history, knowledge of competitors' bidding habits, learning capabilities. Implementation of such bidding behaviour involves feedback loops, memory and general decisional capabilities. An example of non-truthful such strategy is Generalised Second Price (GSP) auction (Russell & Norvig, 1995; Varian, 2007) in which Vickrey and VCG are extended to bidding for multiple objects in multiple iterations. Another example is the greedy strategy (Cary, et al., 2007) which considers that recent past is the best prediction for future. As a result, bidders bid in the current run exactly what they bid in the immediately previous run.

In terms of the proposed RRA SoM agent framework, any of these approaches can be successfully implemented, depending on the specific purpose the agent if built for. However, one should bear in mind that internal agencies of a RRA SoM agent are only components of the SoM agent *internal society*, as envisioned by Minsky, not stand-alone agents with full decisional capabilities and autonomy. They only act within the boundaries of SoM agent's mind, hence they are from a point of view *non-intelligent* components. Only if taken together they can create an overall intelligent behaviour as an SoM agent. Thus, from the point of view of this thesis and the proposed RRA SoM agent, bidding process should not involve complex bidding strategies for it requires a memory of their own, or other complex mechanisms.

Since the RRA SoM agent chooses the action course proposed by the agency with highest strength, the internal competition falls in the category of truthful bidding strategies. Thus, internal agencies should bid with their true value of *strength*. Also, the winning action course belongs to the winning agency, hence from a strategy point of view the recommended implementation methodology is the classic English auction. This particular methodology will be further used throughout the thesis in the following chapters. However, it should be once more emphasised that the RRA SoM architecture also supports other bidding strategies, depending on the specific task SoM agent must accomplish.

3.4. RRA SoM architecture – a proposed JAVA class structure for general implementation

Based on the general implementation methodology described in previous sections of this chapter, this section presents a possible JAVA class structure for the proposed RRA SoM agent. The full code for JAVA classes declaration is presented at the end of thesis in Appendix 1: RRA SoM agent - JAVA class structure**.** The appendix offers a detailed view on how the class structure should be implemented, starting from the general internal dynamics of RRA SoM agent and according to the implementation principles discussed in this chapter.

A summary of the JAVA code is shown in Figure 3.12 as an UML 2.0 class structure diagram. The diagram presents the main structural approach on RRA

SoM agent, emphasising on two aspects implemented as class *SoMagent* and class *Environment*.

Class *Environment* defines the environment in which RRA SoM agents are placed in order to accomplish their specific tasks. It can be seen on the diagram – that also results from the architectural description presented in this chapter – that RRA SoM agents do not possess an explicit direct interaction mechanism. Instead, the interaction is provided by changes produced to environment by agents' actions, hence an implicit indirect interaction is implemented.

Class *SoMagent* defines one individual RRA SoM agent and captures agent's internal structure through its members (methods and properties). It also shows its internal dynamics through member's interdependence and the indirect interaction with other agents through its relation to **class *Environment***.

Interface *Agency* and classes **Agency*. Interface *Agency* describes a generic *Agency* entity which can be *implemented* by any number of concrete **Agency* classes (agency types). In RRA SoM case three classes are defined: *AffectiveAgency*, *RationalAgency*, and *ReactiveAgency*. (*Agency* can be also seen as an abstract class if needed. In this case classes **Agency* will extend the abstract class *Agency*. However, if an abstract class is used instead of an interface, then all relevant internal methods must be declared as *overrideable* and the concrete corresponding classes should be changed accordingly. For simplicity reasons, the interface was preferred.)

Agency* objects take the appropriate contextual (from *Environment* through *SoMagent* properties) and non-contextual inputs from ***SoMagent* object, update their states, calculate *strengths* and propose sets of actions. Actions selected using *Kline* object are passed to *Bidding* object.

Class *Kline* acts on classes **Agency* for selecting the applicable actions for current situation. In terms of process flow, K-Line selection is different for each agency. Thus, from an implementation point of view each type of selection should be a method of the corresponding **Agency* class. However, conceptually the K-Line selection is a stand-alone key element in SoM architecture. For this reason *Kline* appears in JAVA implementation as a different class, but its methods are *static* and accessible by all **Agency* objects when needed.

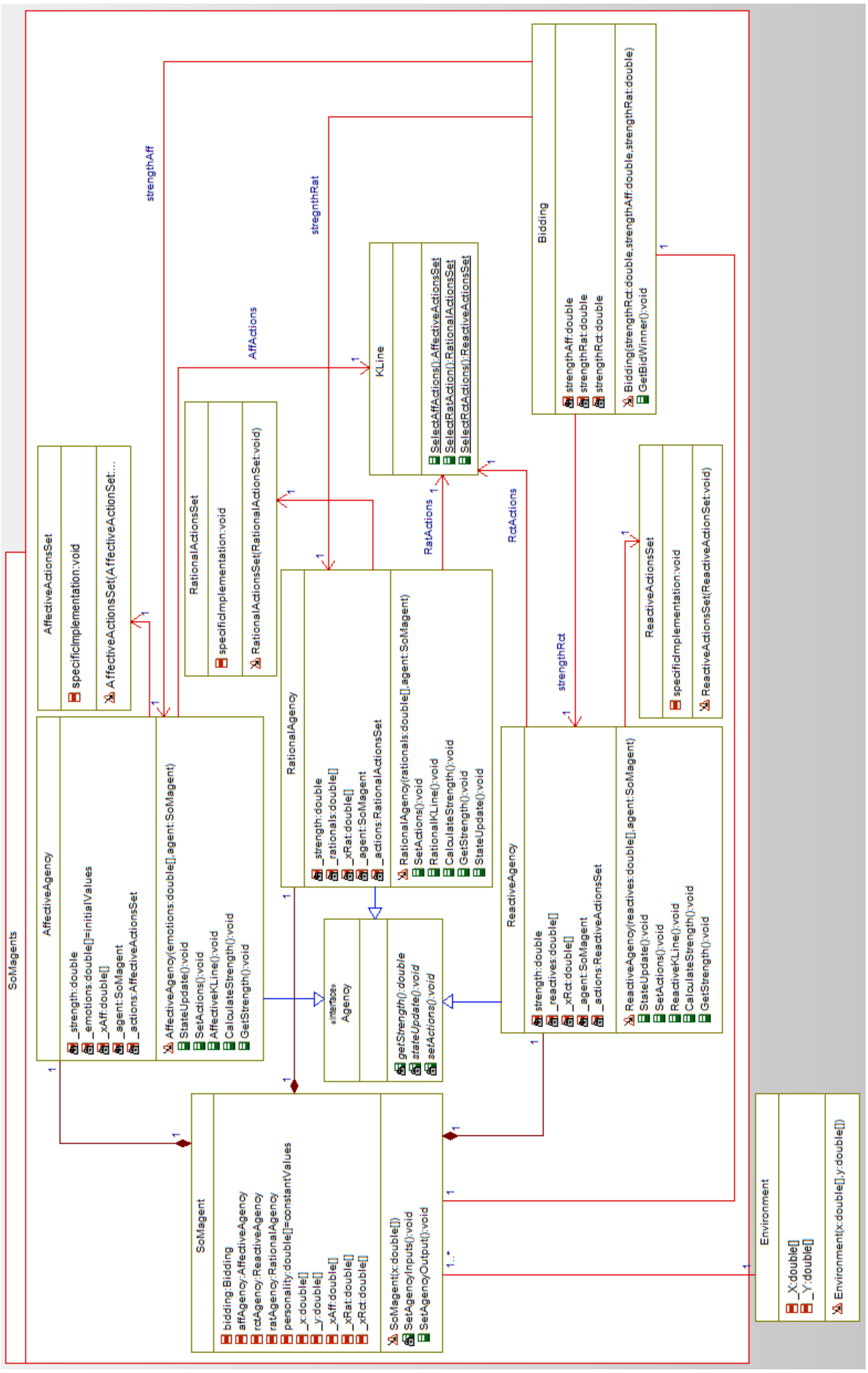


Figure 3.12 RRA SoM agent – an UML 2.0 class structure diagram

Class *Bidding* implements the non-compromise principle, i.e. it implements the bidding strategy and decides the winner. A *Bidding* object takes as inputs *action-strength* pairs from *AffectiveAgency*, *RationalAgency*, and *ReactiveAgency* objects and passes the result (winning action) to *Environment* object through *SoMagent* properties.

3.5. Discussion

This chapter introduced a plausible non-hierarchical hybrid agent architecture based on Minsky's "Society on Mind" metaphor on human mind and cognition. First, it provided a detailed insight on Minsky's approach and its possible applications. Then, a computational instantiation of Minsky's ideas was presented as a general architectural schema of a SoM complex cognitive agent. The general architectural schema was further explained in more detail, with emphasis on internal dynamics of SoM agent's competing sub-agencies and their implementation methodology. This was done using a specific setup with three internal agencies standing for the main agent architectural approaches (and cognitive capabilities): affective, rational and reactive. In the end a potential JAVA class implementation was presented, together with the corresponding UML 2.0 class structure diagram.

Arguably, the proposed agent architecture is able to mitigate some of the drawbacks of existing hybrid cognitive agents, such as the problem-dependency and the logical and computational complexity. In the first place, the general SoM agent architecture eliminates the complexity of layered hierarchical constructions, by replacing complex layer interaction and hierarchical control framework with a simple competition-based process. On the other hand, internal dynamics provided by the architectural design could provide good portability among various problems and fields of activity.

With regard to the research questions of this thesis, this chapter answered the research question number 1.

Chapter 4. An SoM driver agent. Implementation and evaluation

In this chapter the general SoM agent architecture proposed in previous chapter is implemented and evaluated in a cognitively demanding context: as a car driver agent. A SoM driver agent is implemented starting from the general SoM architecture and tested in various traffic situations. The purpose is to demonstrate that such an implementation can produce a wide range of human behaviours and hence, it can be used successfully in modelling and/or analysing complex human activity contexts.

The chapter is organised in three major sections. First section describes implementation of the SoM driver agent as a particular case of the general SoM architecture. In second section the individual SoM the driver agent is evaluated in various car-following situations in order to record and discuss the internal decision making mechanism and the effect it actuates in traffic conditions. Last section discusses the findings and concludes on the contribution of this chapter.

4.1. An SoM driver agent. Implementation

The RRA SoM architecture proposed in Chapter 3 is instantiated in a cognitively diverse and demanding context: the road traffic behaviour. In general, a human driver is considered to take three types of decisions when in traffic, which in the SoM agent context stand for the three types of agencies – rational, affective and reactive:

- Rational agency: a driver takes rational decisions in order to drive safe, to avoid annoying other drivers and to obey traffic regulations;
- Affective agency: a driver takes decisions based on the current emotional state regardless the safety, the annoyance produced to other drivers, and the traffic regulations;

- Reactive agency: a driver MUST take avoiding actions (full brake, lane change) in order to escape from an immediate danger, such as a collision.

Implementation of the agent using the SoM general framework is described in the following paragraphs. However, before proceeding to a detailed description of the proposed implementation, the schema of general architectural framework is reproduced from Chapter 3 in Figure 4.1, in order to serve as a better visual reference.

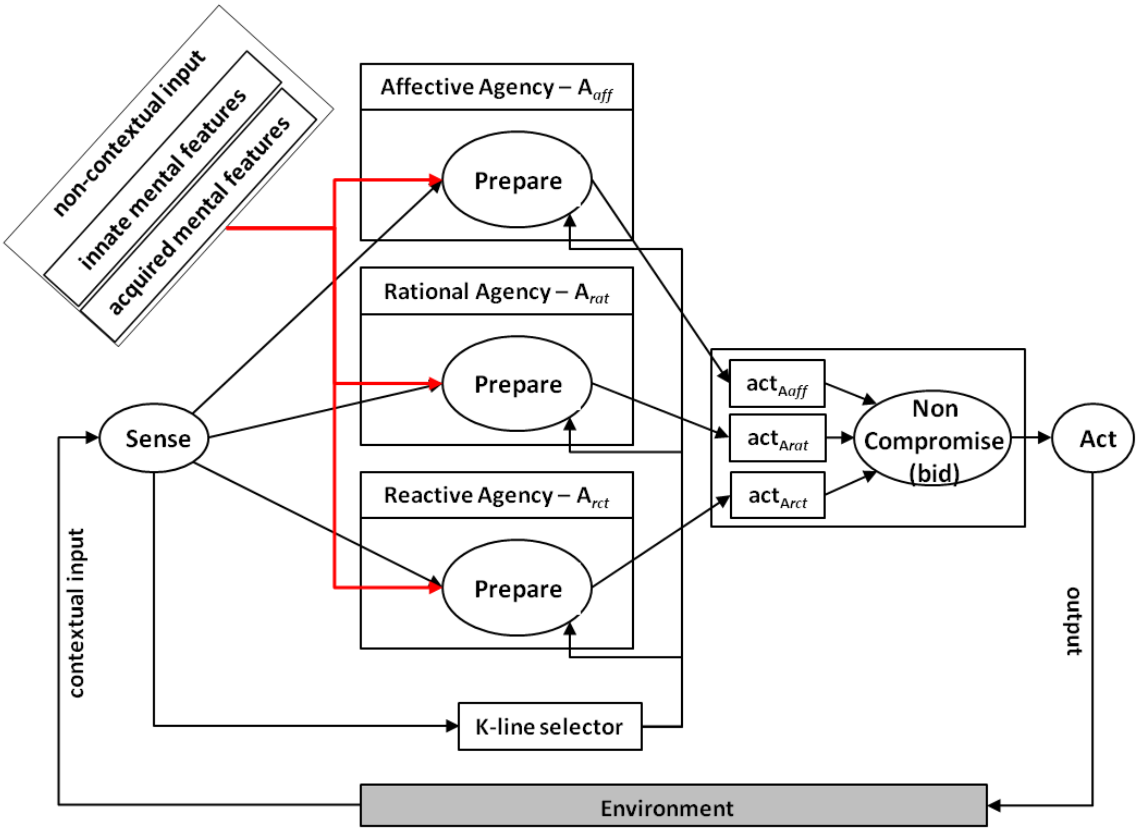


Figure 4.1 Complete RRA SoM architectural scheme with inclusion of innate biases and behavioural propensities.

4.1.1. *Behavioural propensities – innate Personality features*

As discussed in Chapter 3, one of the important aspects of the proposed SoM architectural framework is the capability to handle behavioural propensities, either innate or acquired. It was also noted in the same chapter that acquired behavioural propensities can be included in a potential instantiation of the proposed architecture. However, discussions about this type of propensities may become very broad and for this reason they are not covered in this thesis. Thus, only the innate behavioural features (personality) are considered for the proposed instantiation, the SoM driver agent.

Implementation of personality, as the expression of innate behavioural propensities, is based on Goldberg's (Goldberg, 1990) BIG 5 model of personality, which was addressed in detail both in Chapter 2 and Chapter 3. From the variety of possible implementation approaches presented in Chapter 3, the finite real interval method was chosen.

The five bipolar personality factors – Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (OCEAN) – are represented as a 5-dimensional tuple $P(P_O, P_C, P_E, P_{\boldsymbol{A}}, P_N)$, with $P_i \epsilon(-\infty; \infty)$, where P_i are the personality factors. In order to obtain a computationally usable representation, values of personality factors P_i are limited to the finite real interval $P_i \epsilon[-P_{max}; P_{max}]$, where P_{max}=1.

Values of personality traits/factors are considered system constants and are assigned to the agent at the beginning of simulation. They are fed to internal agencies during the simulation together with variable contextual information coming from sensors and they participate in the state update process implemented in each agency.

4.1.2. *The affective agency*

The affective system assumed in this implementation uses the dimensional approach on emotions, thoroughly discussed in Chapter 2. Dimensional approach assumes several fundamental independent emotional directions, bipolar and continuous, which represent the variety of human emotional responses. Studies in transportation psychology and behaviour (Grimm, Kroschel, Harris, et al., 2007) found that the space of emotions applicable to drivers consists of three independent emotions: happiness (E_H), drowsiness (E_S) and anger (E_A).

Figure 4.2 shows the influence diagram of action formation process and explains the transfer of driver's affective features into traffic-related actions. The influence of these features on traffic behaviour was presented in detail in Chapter

2, Section 2.2. The influence diagram was generated starting from previous studies of Clarke and Robertson (S. Clarke & Robertson, 2005), Lajunen (Lajunen, 2001) and Jovanovic (Jovanovic, et al., 2011) for personality, and those of Fiedler and Bless (Fiedler & Bless, 2001), Shinar (Shinar, 1998) and Grimm (Grimm, Kroschel, Harris, et al., 2007) for emotions. In terms of the potential traffic-related actions, from the virtually infinite number of possible actions of a driver in traffic conditions, the list was reduced to those of maximum importance (Grimm, Kroschel, Harris, et al., 2007; Jovanovic, et al., 2011; Sarkar, et al., 2000): speeding, ignoring priority, tailgating and cutting.

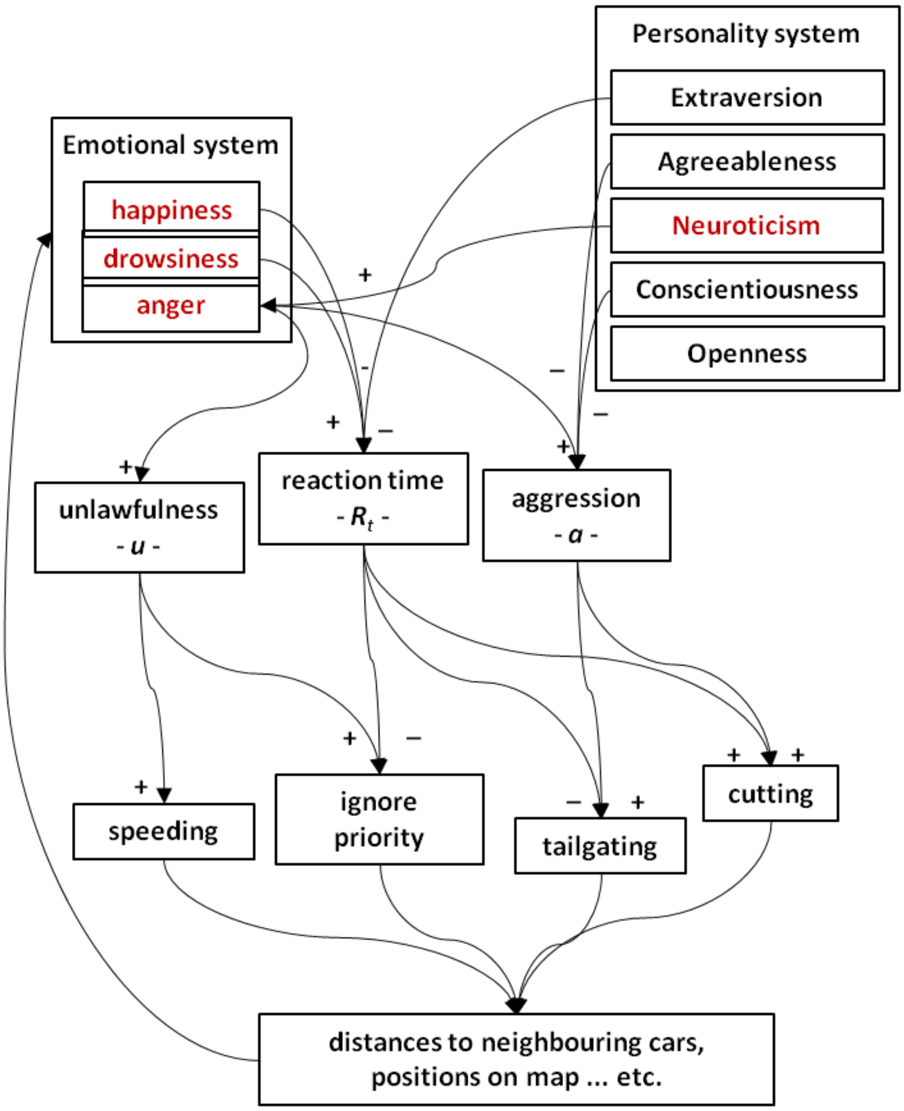


Figure 4.2 Influence of affective features on driving actions: influence diagram

On the diagram, the long term innate personality traits and the short term context-generated emotional states interact. This interaction forms the mental state which raises the appropriate action tendency values to certain levels that could trigger various affect-based actions.

From the implementation point of view emotional space is represented as a 3-dimensional tuple $E(E_H, E_S, E_{\boldsymbol{A}})$, with $E_i \epsilon(-\infty; \infty)$, where E_i are the primitive emotions. Computationally, the problem generated by the infinite emotional intervals was solved following the same approach used for personality: the infinite values of emotions E_i were conveniently limited to finite real intervals $E_i \epsilon[-E_{max}; E_{max}]$, where E_{max}=1.

Figure 4.3 shows the internal dynamics of the SoM affective agency as an adaptation to SoM driver agent implementation of the general influence diagram presented above. It should also be noted that Figure 4.2 presents a general influence diagram, which explains the action formation mechanism, i.e. the transfer of mental states into decision and action. Figure 4.3 presents the actual implementation of the emotional update, which refers to the specific inputs and outputs considered for the SoM driver agent. In other words, Figure 4.2shows that the emotional states are updated, while Figure 4.3 also shows how they are updated.

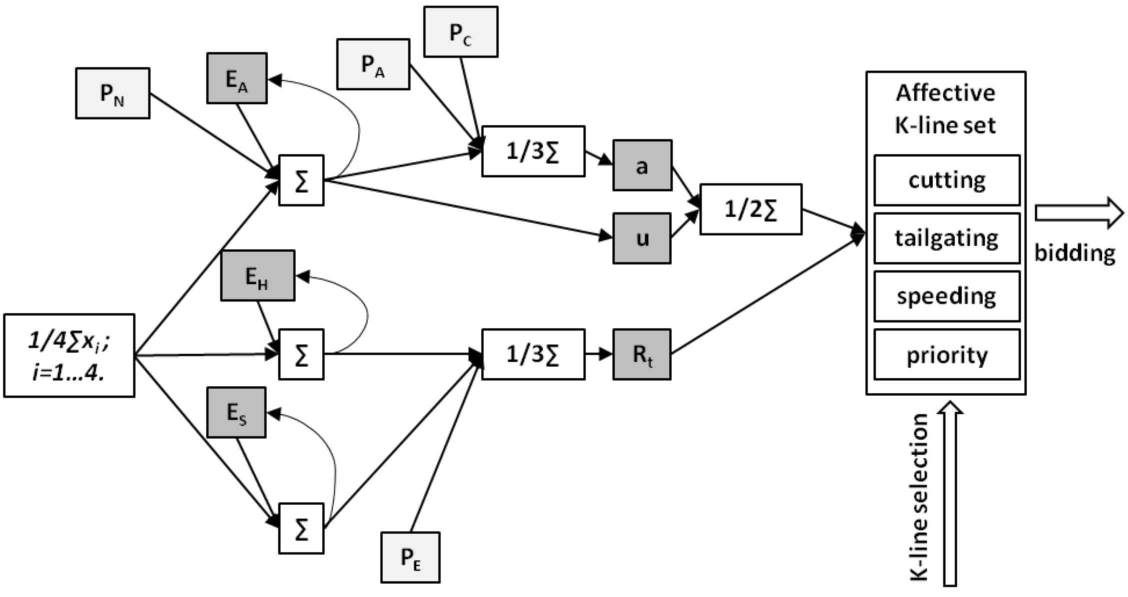


Figure 4.3 The three stage action preparation diagram

Sense:

At simulation time-step t_n the affective agency prepares for potential actions at t_{n+1} based on three inputs: distances to neighbours at t_n, unchangeable values of personality traits, and values of emotions at t_{n-1}. Personality and emotional inputs are situated in the interval [*-1;1*], but distances to neighbours are expressed in metres. These distances can vary widely according to instantaneous positions of vehicles on the road. This generates the need to scale the distance inputs taken from physical environment into values usable in conjunction with the values of personality factors and emotions. Yet, a simple normalisation process would not be sufficient since the intervals of variation are not constant and so, their limits are not fixed values throughout the simulation.

In order to cope with such an issue, deviation of real distance to neighbours from a minimum safe distance is used instead of using the actual distance. In the following paragraphs this process is explained in detail both graphically and analytically.

Figure 4.4 (a) shows the most general traffic pattern (motif) with the current vehicle of interest *c* running from left to right on a double lane road and situated between four neighbours. Distances to these neighbours – d_{1r}, d_{2r}, d_{3r}, d_{4r} – vary in time and for this reason they cannot be placed in a fixed variation range/interval.

An approach to overcome this difficulty is presented in Figure 4.4 (b) and (c), in which the scaling procedure is presented for the distance regarding current vehicle *c* and front neighbour 1. The procedure for all other neighbours is similar.

In this car-following situation, an additional point *s* is considered as in the Collision-Avoidance car-following models presented in Subsection 4.1.3 of this chapter. This point is situated at the minimum safety distance d_{1s} to leading vehicle. This is the distance within which a collision is unavoidable if driver of the vehicle in front acts unpredictably, and given an assumed maximum deceleration capability. Distance d_{1s} is calculated according to Gipps version of Collision-Avoidance models, as in equation (4.14).

However, in real traffic situations the distance between current and leading vehicle is not always d_{1s} as in the ideal Collision-Avoidance case, but it can be either below or above the safe distance. This generates a measurable deviation Δ_1, equation (4.1), and subsequently a normalised deviation x_1 which can fall into two cases.

$$\Delta_1 = d_{1s} - d_{1r} \tag{4.1}$$

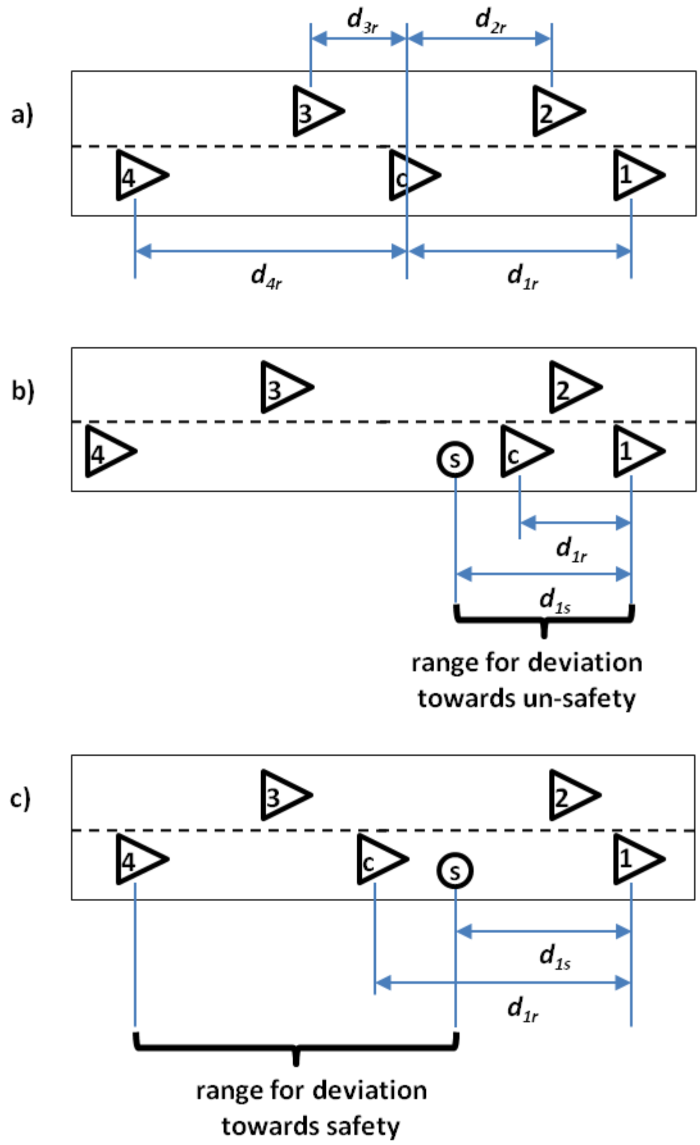


Figure 4.4 Input scaling process: a) unscaled inputs – distances from current vehicle (c) to neighbours; b) relation between current vehicle and neighbour 1 – deviation from optimal/safe distance towards unsafety; c) relation between current vehicle and neighbour 1 – deviation from optimal/safe distance towards safety.

First case, Figure 4.4 (b), is when real distance d_{1r} is below the safe distance ($d_{1r}<d_{1s}$). In this case deviation from ideal point will be towards unsafety. The

range of this deviation is between zero – current vehicle keeps a safe distance $d_{1s}=d_{1r}$ to leading vehicle, and a maximum deviation of d_{1s} – current vehicle follows the leading vehicle at a hypothetical $d_{1r}=0$. Thus, deviation from safe distance can vary in the interval $(0;d_{1s})$. The value used as denominator for normalised deviation x is d_{1s}, equation (4.2), with resultant values of x_1 situated in interval $(0;1)$.

$$\boldsymbol{x_1} = \frac{\Delta}{\boldsymbol{d_{1s}}} = \frac{\boldsymbol{d_{1s}} - \boldsymbol{d_{1r}}}{\boldsymbol{d_{1s}}} \tag{4.2}$$

Second case, Figure 4.4 (s), is when real distance d_{1r} is above the safe distance $(d_{1r}>d_{1s})$. In this case deviation Δ from ideal point will be towards even more safety. The deviation range will be between zero deviation – when $d_{1s}=d_{1r}$, and a maximum hypothetical deviation of d_{4r} – when real following distance $d_{1r}=d_{4r}+d_{1s}$. Thus, deviation from safe distance can vary in the interval $(-d_{4r};0)$. The denominator used for calculating normalised value x is d_{4r}, equation (4.3), with resultant values of x_1 in interval $(-1;0)$.

$$\boldsymbol{x_1} = \frac{\Delta}{\boldsymbol{d_{4r}}} = \frac{\boldsymbol{d_{1s}} - \boldsymbol{d_{1r}}}{\boldsymbol{d_{4r}}} \tag{4.3}$$

Overall, the input scaling process for relation between current vehicle c and neighbour 1 can be described as in equation (4.4). The resultant scaled input $x_1 \in (-1;1)$ can now be used in conjunction with the other relevant variable and constants in the process of updating the state of affective agency.

$$\boldsymbol{x_1} = \begin{cases} \dfrac{\Delta}{\boldsymbol{d_{1s}}}; \boldsymbol{if}\ \boldsymbol{d_{1s}} > \boldsymbol{d_{1r}} \\ \dfrac{\Delta}{\boldsymbol{d_{4r}}}; \boldsymbol{if}\ \boldsymbol{d_{1s}} < \boldsymbol{d_{1r}} \end{cases} \tag{4.4}$$

Prepare:

The 'prepare' stage consists of two main elements: the emotional update function and the strength update function.

Emotion update function was implemented starting from the intensity-decay approach described in Cathexis model (Velásquez, 1997), as explained earlier in the thesis, in Chapter 3. Velasquez assumes that, unless there is some inhibitory or excitatory input, the intensity of an emotion decays according to a predefined decay function in every update cycle, as in equation (3.2) presented in Chapter 3. For a better visual reference, equation (3.2) is reproduced below as equation (4.5):

$$\boldsymbol{E_{it}} = \boldsymbol{f}\left\{\boldsymbol{D}(\boldsymbol{E_{it-1}}) + \sum_{\boldsymbol{k}} \boldsymbol{L_{ki}} + \sum_{\boldsymbol{l}} \boldsymbol{G_{li}}\, \boldsymbol{E_{lt}} - \sum_{\boldsymbol{m}} \boldsymbol{H_{mi}} \boldsymbol{E_{mt}}\right\} \tag{4.5}$$

where E_{it} is the intensity of emotion i at time t, $D()$ is he decay function of emotion i, L_{ki} is the elicitor k of emotion i, G_{li} is the excitatory gain that emotion l applies on emotion i, H_{mi} is the inhibitory gain that emotion m applies to emotion i and f is the function that places the emotion i between 0 and its maximum value.

This equation is further adapted to fit the dimensional approach on emotions in traffic conditions proposed by Grimm and colleagues (Grimm, Kroschel, Harris, et al., 2007). It should be noted that Grimm considers the emotions to be independent. Thus, the inhibitory and excitatory influences emotions have on each other are left out of the calculation. Only a single functional entity that includes contribution of both decay function and elicitors is considered, as in (4.6). The three emotions considered in this implementation are: happiness, drowsiness and anger (Grimm, Kroschel, Harris, et al., 2007). They increase or decrease at each time step as in (4.6), where x_i are the scaled inputs corresponding to instantaneous distance-speed relation to the four neighbours. If the resultant values of emotions exceed interval *(-1;1)* then flooring and ceiling functions are activated to limit the peak values to interval's extremes *-1* and *1*.

From an intuitive perspective equation (4.6) implements the following. Happiness increases when the real distance is *higher than* the minimal desired value and decreases as it goes smaller. Drowsiness (loss of attention, increased reaction time) increases when the real distance is *much higher than* the minimal desired value, and decreases when it goes back to acceptable values. The *much higher* concept is implemented in the equation by dividing the distance to neighbours by two. Thus, the contribution of the second term of E_D update function becomes significant only for distances to neighbours which are large enough to produce drowsiness. Anger, increases when the real distance is *lower than* the minimum safe value and decreases as it goes back to higher values.

In the above paragraph expressions such as *higher than*, *much higher than* and *lower than* were used for describing the possible situations in which two vehicles can be on the road in terms of the distance between them. Even if the expressions may appear vague, their representation in equation (4.6) actually covers the whole range of situations used in car-following agent models, such as those proposed by Panwai and Dia (Panwai & Dia, 2007). In their approach driving conditions with respect to two vehicles following each-other fall into five categories: *free driving*, *approaching*, *following I*, *following II* and *danger*. These categories consider in a discrete manner the distance ranges from larger than 60 m (free driving) to lower than 10 m (danger) and neglect the influence of speed on the safe following distance. Arguably, equation (4.6), provides a better design approach. The emotional sates are updated by the relative distance between cars, calculated with regard to the speed-related minimal safe distance provided by collision-avoidance model (4.14). As a result, a continuous and speed dependant approach on the driving situations is implemented.

$$\begin{cases} \boldsymbol{E_H(t_n) = E_H(t_{n-1}) + \frac{1}{4}\sum_{i=1}^{4} x_i(t_n)} \\ \boldsymbol{E_D(t_n) = E_D(t_{n-1}) + \frac{1}{4}\sum_{i=1}^{4} x_i(t_n)/2} \\ \boldsymbol{E_A(t_n) = E_A(t_{n-1}) + P_N + \frac{1}{4}\sum_{i=1}^{4} x_i(t_n)} \end{cases} \tag{4.6}$$

Strength update function. After the emotional states are updated they are used together with personality traits and scaled contextual inputs for transferring the updated affective state into a *strength* level. The strength value will further participate in the bidding process included in *non-compromise* stage of decision. The update function of the *strength* is presented in (4.7).

$$\boldsymbol{aff_{strength}(t_n)} = \frac{\boldsymbol{a(t_n) + u(t_n)}}{\boldsymbol{2}} \tag{4.7}$$

where $\boldsymbol{a}$ and $\boldsymbol{u}$ are aggression and unlawfulness levels, respectively, calculated as in equations (4.8) and (4.9). Affective strength is also in the interval *(-1;1)* and for this reason is further shifted up with one unit and normalised, so it eventually falls into the interval *(0;1)* which is used in the bidding process.

$$\boldsymbol{a(t_n)} = \frac{\boldsymbol{E_A(t_n) - P_A - P_C}}{\boldsymbol{3}} \tag{4.8}$$

$$\boldsymbol{u(t_n) = E_A(t_n)} \tag{4.9}$$

K-Line selection, and actuation:

Apart from being used in the bidding process, the value of affective strength also shows the magnitude of proposed actions and it is used for actuation in various physical contexts identified by the K-line selector. These physical contexts are called in this thesis “traffic motifs”. They are briefly presented in Table 4.1, and are explained and treated separately in the second part of this chapter.

However, all proposed actions regardless the motif they belong to, can be actuated by two output variables: speed, and lane change (if applicable).

The vehicle is proposed to accelerate/decelerate with a fraction of maximum acceleration/deceleration proportional to affective strength. Its new speed is then calculated taking into account the proposed acceleration over a period of time with duration of a simulation time-step t_s (4.10). Actuation of speed accounts in physical environment for affective actions of speeding, tailgating and ignoring priority.

	Motif	Description	Lane setting
1	C	current vehicle	single lane
2	C1	current vehicle and neighbour 1	single lane
3	C4	current vehicle and neighbour 4	single lane
4	C14	current vehicle and neighbours 1 and 4	single lane
5	C12	current vehicle and neighbours 1 and 2	double lane
6	C34	current vehicle and neighbours 3 and 4	double lane
7	C124	current vehicle and neighbours 1, 2 and 4	double lane
8	C134	current vehicle and neighbours 1, 2 and 4	double lane
9	C1234	current vehicle with all neighbours	double lane

Table 4.1 Traffic motifs

Lane change is also proposed based on the affective strength, and is modelled as a threshold function, (4.11), in which a lane change is triggered if the affective strength is greater than 0. Actuation of lane change accounts in physical environment for actions of cutting (and weaving, which is not treated in this thesis).

$$\boldsymbol{v_{aff}(t_n)} = \begin{cases} \boldsymbol{v_{aff}(t_{n-1}) + t_S \cdot aff_{str}(t_n) \cdot acc_{max}; if\ aff_{str}} > 0 \\ \boldsymbol{v_{aff}(t_{n-1}) - t_S \cdot aff_{str}(t_n) \cdot dec_{max}; if\ aff_{str}} < 0 \end{cases} \tag{4.10}$$

$$\boldsymbol{laneCh_{aff}(t_n)} = \begin{cases} \boldsymbol{1; if\ aff_{str}(t_n)} > 0 \\ \boldsymbol{0;\ if\ aff_{str}(t_n)} < 0 \end{cases} \tag{4.11}$$

4.1.3. *The rational agency*

As explained in the introductory part of this section, the rational nature of a driver is always concerned with safety, politeness and rule obeying. This agency is targeting optimality from the safety point of view, by enforcing safe following distances correlated with vehicle speed. An appropriate way to implement this behavioural pattern is to use one of the collision-avoidance car-following models, which generate a prescribed safe behaviour based on Newtonian equations of motion. Thorough reviews of various equation-based car-following and lane-changing models, including collision-avoidance models can be found in (Brackstone & McDonald, 1999), (Kesting, Treiber, & Helbing, 2007) and (Weng & Wu, 2001).

Most of the microscopic single-lane traffic models describe the motion of a vehicle *n* as a function of its own speed v_n, the distance to the front vehicle x_n and the relative velocity between them Δv_n. They implement vehicle behaviour using the classical equations of motion by computing instantaneous accelerations, velocities or distances for each vehicle at the current step of the simulation. In the most general case acceleration of the vehicle *n* can be expressed (Kesting, et al., 2007) as in (4.12).

$$\boldsymbol{a_n \equiv \frac{dv_n}{dt} = a(x_n, v_n, \Delta v_n)} \tag{4.12}$$

The most important models following this approach, such as Gazis-Herman-Rothery (GHR), Optimal Velocity, Collision Avoidance (CA), Linear (Helly), Action Point or Fuzzy-logic-based, are thoroughly reviewed in (Brackstone & McDonald, 1999; Kesting, et al., 2007; Weng & Wu, 2001). In this thesis driver behaviour is the essential concern. This makes GHR and CA models appropriate for use, because they offer the possibility to upgrade the standard equations of motion with desired behavioural rules requested by the mental state model. The GHR model computes the instantaneous acceleration as in (4.13) and describes a stimulus-response type of function like in (4.12).

$$\boldsymbol{a_n(t) = c v_n^m(t) \frac{\Delta v(t-T)}{\Delta x^l(t-T)}} \tag{4.13}$$

where the acceleration a_n of vehicle *n* at time *t* is proportional to its own velocity v_n at time *t* and to its relative spacing and speed Δx and Δv to the next (leading) vehicle at an earlier time *t*-*T* where, *T* is the reaction time of the driver; *m*, *l*, *c* are model's empirically determined constants.

The CA model does not explicitly describe a stimulus-response function, though it implicitly contains it. Instead, it assumes a minimum safety distance within which a collision is unavoidable if the driver of the vehicle in front acts unpredictably, and given an assumed maximum deceleration capability, as in (4.14):

$$\Delta \boldsymbol{x}(\boldsymbol{t_n}-\boldsymbol{T}) = \boldsymbol{\alpha v^2_{fol}}(\boldsymbol{t_n}-\boldsymbol{T}) + \boldsymbol{\beta_l v^2_{lead}}(\boldsymbol{t_n}) + \boldsymbol{\beta v_{lead}}(\boldsymbol{t_n}) + \boldsymbol{b_0} \tag{4.14}$$

where Δx is the minimum safe distance between leading and following vehicles, v_{lead} is the speed of leading vehicle, v_{fol} is the speed of following vehicle. Empirical parameters of the Collision-Avoidance model are those presented in (Brackstone & McDonald, 1999): unique reaction time T=0.75s, α=0.00028, β_l= -0.0084, β=0.784 and b_0=4.1.

The GHR model, through its explicit implementation of a stimulus-response function, seems to be more attractive for a behavioural approach. However, the CA approach was chosen because it is more appropriate for usage in conjunction with one of the existing gap-acceptance lane-changing models (Kesting, et al., 2007; Laval & Daganzo, 2006) for creating multi-lane traffic models.

Collision-Avoidance models have been successfully used not only in simple car-following situations, but also in multi-lane traffic environments for lane-changing tasks as part of gap-acceptance models. Gap-acceptance models are consistent with their collision-avoidance counterpart from the single-lane car-following models. They assume that a lane-change depends on the existence of an acceptable (safe) gap between the current and the neighbouring vehicles, i.e. following and leading vehicles in the current and target lanes. Depending on the scope and purpose of the investigation not only the front vehicle but any configuration of the neighbouring vehicles can be taken into account (Kesting, et al., 2007). Usage of the collision-avoidance car-following model in a gap-acceptance setup creates an overall collision free traffic model. It also brings the advantage of a compact mathematical formulation, since the minimum distance equation described in (4.14) can be successfully used simultaneously for both line-changing and car-following.

In this study Gipps model (Wilson, 2001) is used, an enhanced version of the classic CA approach which provided best results in car-following modelling (Panwai & Dia, 2005) both analytically and numerically (traffic simulators). Gipps model is presented in equation (4.15):

$$\begin{aligned} \boldsymbol{v_n} &= \boldsymbol{min}\left[\boldsymbol{v_n(t)} + \boldsymbol{2.5A_n\tau}\left(\boldsymbol{1} - \frac{\boldsymbol{v_n(t)}}{\boldsymbol{V_n^{max}}}\right)\left(\boldsymbol{0.025} + \frac{\boldsymbol{v_n(t)}}{\boldsymbol{V_n^{max}}}\right)^{1/2} ;\ -\boldsymbol{B_n}\left(\frac{\boldsymbol{\tau}}{\boldsymbol{2}} + \boldsymbol{\theta}\right)\right. \\ &\left. + \sqrt{\left[\boldsymbol{B_n^2}\left(\frac{\boldsymbol{\tau}}{\boldsymbol{2}} + \boldsymbol{\theta}\right)^2 + \boldsymbol{B_n}\left\{\boldsymbol{2}\{\boldsymbol{x_{n-1}(t)} - \boldsymbol{x_n(t)} - \boldsymbol{S_{n-1}}\} - \boldsymbol{\tau v_n(t)} + \frac{\boldsymbol{v_{n-1}(t)^2}}{\boldsymbol{B_{n-1}}}\right\}\right]}\right] \end{aligned} \tag{4.15}$$

where *n* and *n-1* refer to following and leading vehicle respectively and *v* represents speed, , *A* and *B* represent maximum acceleration and deceleration capability, V_n^{max} is the maximum speed at which following driver wishes to travel, τ is the reaction time and $\theta = \tau /2$ is a safety margin taken by drivers in normal conditions according to Gipps.

Sense:

At simulation time-step $\boldsymbol{t_n}$ the prescribed-for-safety agency prepares for potential actions at $\boldsymbol{t_{n+1}}$ based on the distances to all four possible neighbours at $\boldsymbol{t_n}$ (neighbours 1, 2, 3 and 4) and also using the values of personality traits. An input scaling process identical with that used for affective agency is used for making physical world inputs compatible with *(-1;1)* interval used by personality traits values.

Prepare:

Update rational entities. According to equation (4.6) the rational agency computes the minimum distance d_{1s} that current vehicle *c* must keep from leading vehicle (neighbour 1) given the current speeds of the two vehicles (avoid tailgating). This is further used in Gipps model to adjust the speed of current car *c* with regard to the safety distance, as in equation (4.15).

In the same time, depending on the physical situation on the road, the possibility of changing lane is also checked by using equation (4.6) for current vehicle with neighbours 2 and 3 respectively. In order to avoid cutting (change lane with unsafe gap from the following car) a role inversion is used in Gipps equation: current vehicle *c* is considered leading vehicle while neighbour 3 is considered following vehicle. In order to avoid tailgating (change lane with unsafe gap to the car in front) a second role inversion is used in Gipps equation: current vehicle *c* is considered a following vehicle while neighbour 2 considered a leading vehicle.

Thus the output of the prescribed-for-safety agency is a potential set of actions which, depending on the physical situation on the road, consists of the current vehicle's speed at next time step (t_{n+1}) and the acceptance or denial of lane changing action.

Strength update function is presented in equation (4.16). The rational agency mainly acts in a prescribed manner based on a car-following model. The strength of this agency can be assumed to depend only on the leading car, neighbour 1, and also on the values of personality traits. Equation (4.16) describes the following: the need of rationality increases as the following distance falls below the safe distance and decreases otherwise. Also, rationality is expected to increase as personality traits are on the positive side and decrease otherwise.

$$\boldsymbol{rat_{str}(t_n) = x_1(t_n) + \frac{P_C + P_A - P_N}{3}} \tag{4.16}$$

K-line selection, and actuation:

K-line selection for rational agency is similar to the motif selection proposed for affective agency. A set of traffic motifs identical with the one shown in Table 4.1 is used for identifying the physical context and choosing the appropriate action.

4.1.4. The reactive agency

R Reactive agency implements driver's physiological reaction (reflex) to extreme danger. In traffic conditions extreme danger is considered when a collision with the vehicle in front is unavoidable. Reaction is based on the following elements: deceleration capabilities of the current (following) vehicle, speed difference between the two vehicles, and distance between them.

The reactive agency is implemented using a standard Finite State Machine (FSM) approach on reactive architectures, with the state transition table as in Table 4.2. In this table $\Delta t = t_{n+1} - t_n$ is the simulation time-step, and d_M is the maximum deceleration the vehicle is capable of.

Sense:

At simulation time-step t_n the reactive agency prepares for potential actions at t_{n+1} based on the following inputs: own speed at t_n, distance to front neighbours on the same lane and adjacent lane at t_n, neighbour speeds at t_n.

Prepare:

The agency is implemented in a rule-based approach in which a collision possibility is calculated using manipulations of the standard equations of motion as in (4.17):

$$\Delta \boldsymbol{x_{danger}} = \frac{(\boldsymbol{v_{fol}} - \boldsymbol{v_{lead}})^2}{\boldsymbol{d_M}} \tag{4.17}$$

where Δx_{danger} is the danger distance below which a collision cannot be avoided given the maximum deceleration capability d_M of the current (following) vehicle and the speed difference between the two vehicles.

Strength update function. Unlike the other two agencies, the reactive agency has a binary strength update function (4.18), which signals a dangerous situation with regard to the relation between current vehicle *c* and leading vehicle *neighbour 1*.

If the distance d_{1r} between vehicle *c* and neighbour 1 is equal or lower than Δx_{danger} the reactive agency generates maximum strength. This leads to 100% chances of winning the competition with the other agencies. The actions to be actuated in this case are as follows: the agency proposes a lane change while keeping the existing speed if a gap larger than danger gap exists. If such a gap does not exist the agency proposes full deceleration capability while keeping the current lane.

If distance d_{1r} between vehicle *c* and neighbour 1 is higher than Δx_{danger}, danger does not exist; hence reactive agency is not needed. Thus it generates a disqualifying value of strength which leads to 100% chances of losing the competition with other agencies.

$$\boldsymbol{rct_{str}(t_n)} = \begin{cases} \mathbf{1};\ \boldsymbol{if\ d_{1r} \leq \Delta x_{danger}} \\ \mathbf{0};\ \boldsymbol{if\ d_{1r} > \Delta x_{danger}} \end{cases} \tag{4.18}$$

CURRENT STATE (t_n)	INPUT	NEXT STATE (t_{n+1})
Cruise Speed: $v(t_n)$ Lane No = 1	Danger Adjacent lane: free	Cruise speed: $v(t_{n+1}) = v(t_n)$ Lane No = 2
	Danger Adjacent lane: busy	Reduced Speed: $v(t_{n+1}) = v(t_n) + \Delta t \cdot d_M$ Lane No = 1
Cruise Speed: $v(t_n)$ Lane No = 2	Danger Adjacent lane: free	Cruise speed: $v(t_{n+1}) = v(t_n)$ Lane No = 1
	Danger Adjacent lane: busy	Reduced Speed: $v(t_{n+1}) = v(t_n) + \Delta t \cdot d_M$ Lane No = 2

Table 4.2 State transition table for FSM-based reactive agency

4.1.5. *The "non-compromise" stage. Bidding strategy.*

In Chapter 3, Section 3.3 it was explained that the RRA SoM agent chooses the action course using a truthful bidding strategy based on the classical English auction. Internal agencies bid with their true value of *strength* and winner is the agency with highest strength. The RRA SoM driver agent follows the same strategy with the three agencies bidding with values in interval (0;1).

4.2. Evaluation of the SoM driver agent in a traffic behaviour context

The proposed driver agent instantiation of general RRA SoM agent architecture is tested in a traffic behaviour and psychology context. First the SoM driver agent is evaluated in a standard car-following context against Gipps model. The purpose is to demonstrate that the SoM agent can generate decisions that can deviate from the optimality (rationality) assumed by conventional car-following models. Such a behaviour is made possible by the inclusion of behavioural propensities and instantaneous mental states. Then the agent is evaluated in contexts in which all drivers are SoM agents in a set of motifs which account for all possible situations in which a vehicle can be in traffic conditions at a certain moment in time. This second set of simulations intends to bring an insight on agents' internal decision-making process. It investigates the influence of contextual inputs and innate behavioural propensities on activation of affective, rational and reactive decisions actuated in the environment.

4.2.1. Car-following evaluation

The car-following investigation is derived from one of the most used evaluation methods for computational car-following models (Brackstone & McDonald, 1999; Panwai & Dia, 2005, 2007; Wu, Brackstone, & McDonald, 2003). In this approach the vehicle of interest acts based on a built-in car-following model. The vehicle follows for a certain distance on a straight single-lane road segment a leading car which travels with a prescribed speed pattern. Motion of the leading and following cars is then recorded and their distance trajectories, instantaneous speeds and distance headways are investigated in order to evaluate if the model generates the expected car-following behaviour.

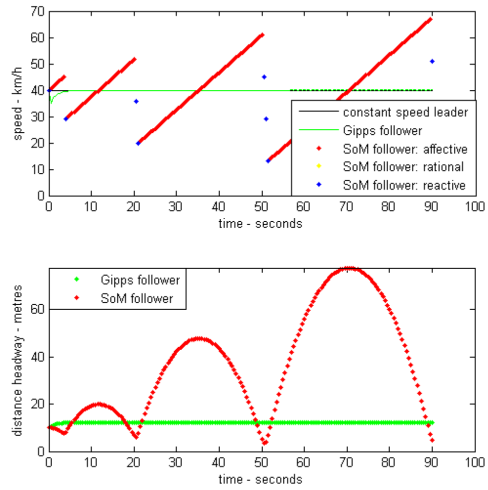
Given that in this case the model to be evaluated is a behavioural one, simulations use a setup with a 1000 m single lane road segment. This is significantly longer than the range of 160 to 200 metre road segments used by most of the car-following model evaluation studies (Panwai & Dia, 2005; Wu, et al., 2003). This setup is used for allowing the mental features of the driver agent to evolve towards the desired decision-making capabilities. The leading vehicle runs at a constant speed of 40 km/h. The following vehicle is controlled by a SoM driver agent that can be initialised with various personality patterns. In addition, the time-step of the simulation t_n was considered t_n=0.5 seconds, in accordance with the findings presented in (Brackstone & McDonald, 1999) for car-following models.

Figure 4.5 shows three situations in which the following SoM driver is initialised with: a – maximum negative personality features *P=P*(*N/A,-1,-1,-1,1*), b – balanced personality *P=P*(*N/A,0,0,0,0*) and c – maximum positive personality features *P=P*(*N/A,-1,-1,-1,1*). In each situation instantaneous speed and distance headway for both SoM driver agent and Gipps follower are computed.

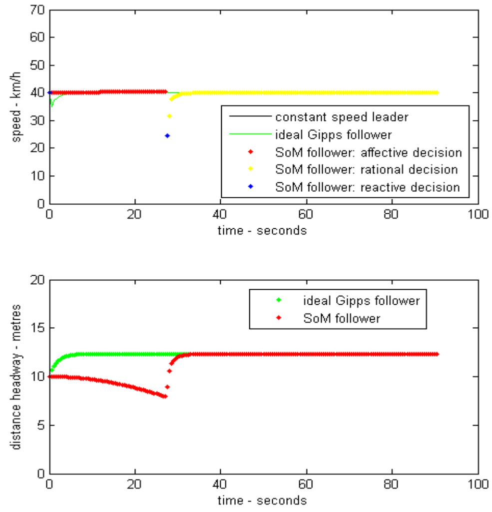
In all cases Gipps follower has the expected rational and safety-speed optimal following behaviour. It rapidly adjusts its speed to a constant value of 40km/h identical to that of leading car.

It also manages to keep a constant minimum distance headway of approximately 12 metres to leading car, which is optimal from a safety-velocity pointy of view, as described by equation (4.15). For the SoM agent follower, decisions actuated in physical environment as instantaneous speed are taken at each simulation step by the current dominant internal agency, being influenced by agent's personality.

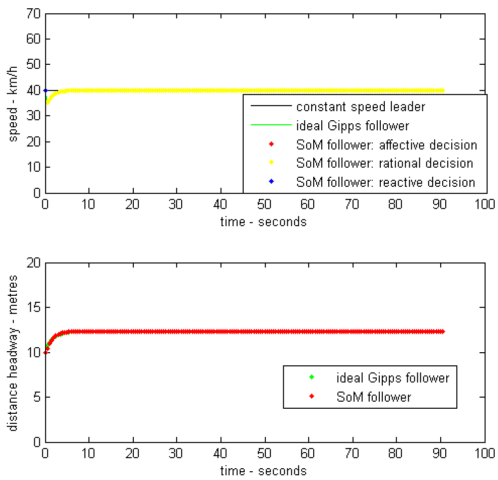
The agent with extreme negative personality traits, depicted in Figure 4.5-a, is highly introverted, neurotic, disagreeable and non-conscientious and has an oscillating following behaviour. Rational decision is absent throughout the simulation with only affective (96.11% of the time) and reactive (3.89%) decisions governing its actions. Repeatedly, the affective agency takes control and dictates rapid speed increase over the safety limit to a dangerous distance to leading car. This fact triggers the reactive agency to take control and actuate reactive breaking in order to avoid collision. Thus, the behaviour of aggressive follower becomes a wave-like motion in which the follower "pressures" the leader by approaching very fast to the danger limit, and then suddenly breaking to avoid collision.



a)



b)



c)

Figure 4.5 Standard car-following setup: instantaneous speed and time headway: a) negative personality traits, b) balanced personality traits, c) positive personality traits

The agent with balanced personality has a behaviour closer to optimality, with the affective agency still becoming active for 29.83% of the time. However, the acceleration rate is lower, the reactive breaking only occurs once (0.055% of time), and the speed at which the reaction is triggered is significantly lower than the previous case. For the rest of 69.61% of time control is taken by the rational agency, which holds the following vehicle on an optimal following pattern similar to that provided by standard Gipps follower.

As expected, the agent with positive personality traits acts at the optimal end of the behavioural spectrum, with rational agency being in control for 100% of the simulation. The resultant car-following behaviour is in this case optimal and identical with that generated using standard Gipps model.

It should be also noted that at a first sight the distance headway in Figure 4.5-a is increasing whenever the speed of affective SoM follower increases. This seems to be in contradiction with Figure 4.5-b where the distance headway is decreasing for affective SoM follower. However, at a more detailed observation, it can be seen in Figure 4.5-a that the distance headway increases while the speed of SoM follower increases but is still lower than the speed of the leader. Thus, it can be said that the increase rate of the distance headway decreases as the follower's speed increases. This happens until the follower's speedreaches leader's speed. Then, as follower's speed continue to increase above the leader's speed, the distance headway decreases. Thus, it can be concluded that results from Figure 4.5-a and Figure 4.5-b are consistent.

From a car-following perspective results showed that driver behaviour varies widely when mental features are included in simulation, producing expected deviations from the rationality/optimality assumed by conventional models (i.e. Gipps model). From an SoM agent perspective, results show that personality biases determine variations in the amount of activation for the affective, rational and reactive agencies.

Thus, it is confirmed a dependency of the activation amount of each agency on agent's personality type. Results also suggest that negative types of personality tend to activate the affective agency more, fact that triggers increased activation of reactive agency as result of unsafety. On the contrary, positive personalities favour activation of rational agency.

A more detailed view of this dependency is shown in Figure 4.6. This summarises results from an extended set of simulations in which agent's personality was swept from most negative to most positive. Results confirm the assumed dependency. However, an equilibrium point between affective (towards unsafety) and rational (towards safety) dominance is not situated in the middle of the two personality extremes. Instead this is rather biased towards the negative side in a range belonging to interval (-0.4;-0.2).

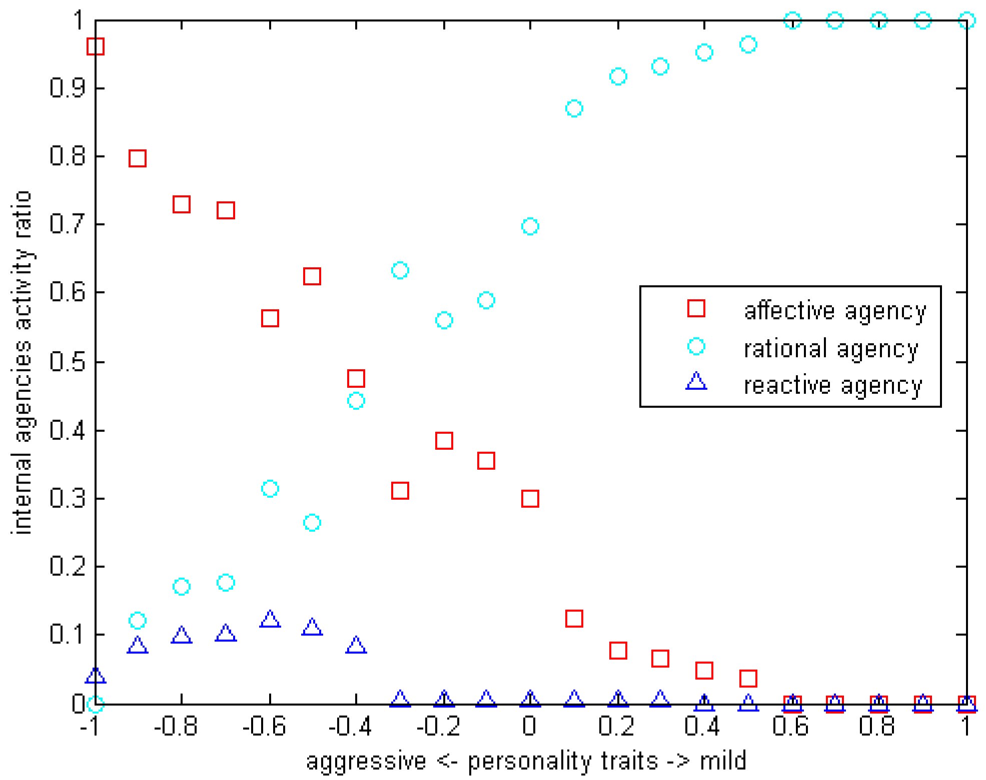


Figure 4.6 Standard car-following setup: agency activation rate (percentage of simulation time) versus personality type

Since the SoM agent takes as inputs distances to all neighbours, a biased activation of internal agencies is not unexpected in this particular situation when only neighbour 1 is taken into account. As described in Subsection 4.1.2 of this chapter activation of affective agency is based on the emotional update function described in equation (4.6). The function is of an additive type and depends on contextual data coming from physical distances to neighbours. Thus, higher increase rates for emotions are expected when inputs from all neighbours contribute to the function. In the case of a single neighbour, only one input from the set of four contributes to emotional update.

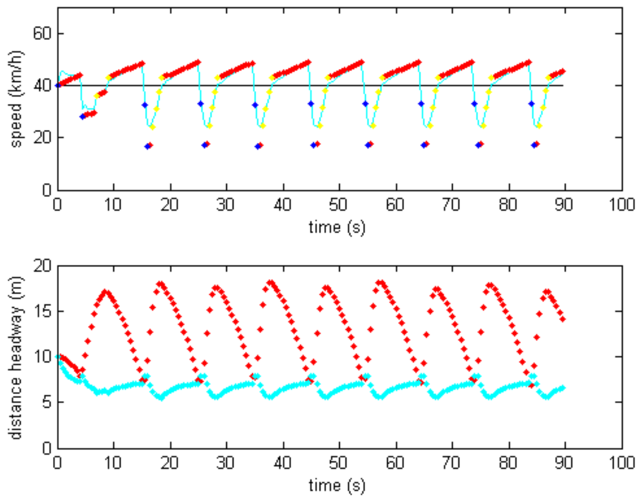
From an intuitive point of view with only one neighbour, there is less potential stress which could raise affective agency's levels of strength, giving more chances for rational decision. As a result, even if the agent has personality traits of a certain level of negativity, there is no pressure coming from the traffic context which can activate affective decisions.

In order to test this dependency on contextual inputs, the standard car-following setup is extended to simulations with two neighbours (1 and 4). These are performed in similar conditions: a 1000 metres single-lane road segment on which vehicle of interest c (the SoM driver agent) follows a constant speed leader (neighbour 1) and is followed by an ideal/rational Gipps follower (neighbour 4).

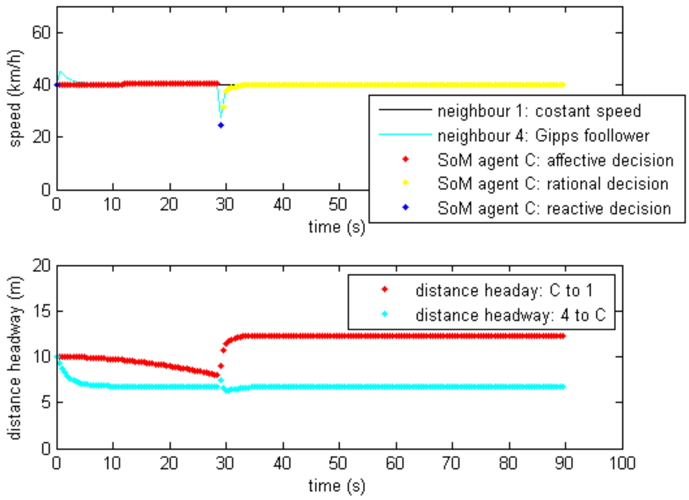
Results presented in Figure 4.7 for agents with negative, balanced and positive personality traits show a similar car-following speed-safety behavioural pattern to those obtained for simple car-following setup. This confirms the resemblance of agent behaviour with expected real-life situations.

From the point of view of internal agency activation, Figure 4.8 also confirms the agency activation variation pattern found in the standard car-following setup. The balanced activation is biased towards negative side of personality traits, though a lower bias was recorded in this case. Potentially this is due to the extra neighbour (second source of stress) whose influence adds to emotion update function and increases the affective agency activation rate for the same personality type.

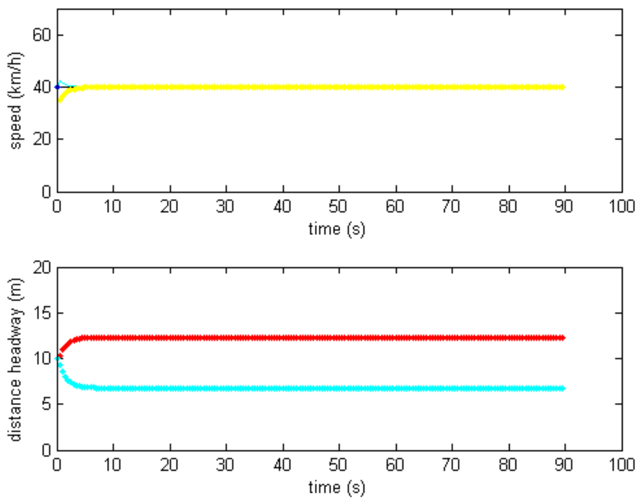
However, a more detailed investigation of internal agencies activation pattern for SoM driver agent is needed in order to draw a clearer conclusion. In order to do that, SoM driver agents must be studied outside the limited context assumed by standard car-following evaluation, i.e. traffic situations with SoM agents only.



a)



b)



c)

Figure 4.7 Modified car-following setup: instantaneous speed and time headway: a) negative personality traits, b) balanced personality traits, c) positive personality traits

In this case both leaders and followers are not ideal (Gipps) but SoM agents with unpredictable personality and context guided driving behaviour. This investigation, with the consequent discussion, is presented in the next section.

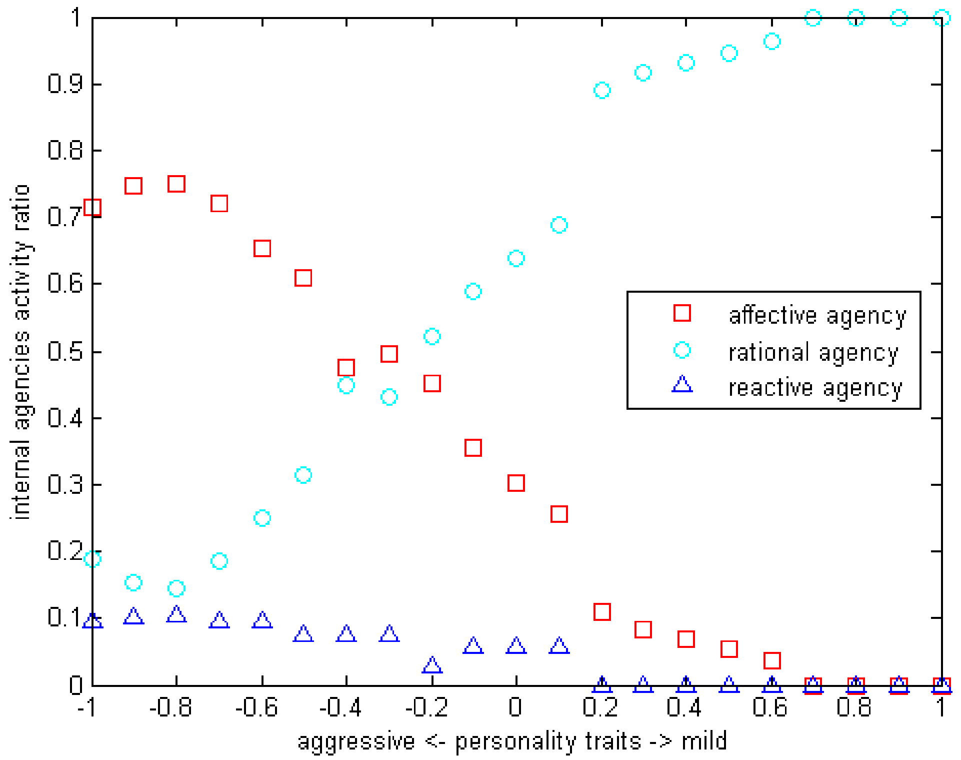


Figure 4.8 Modified car-following setup: agency activation rate (percentage of simulation time) versus personality type

4.2.2. Traffic motifs evaluation

Second set of experiments evaluates the SoM driver agent in traffic situations in which the neighbours are also SoM agents. It was explained in Subsection 4.1.2 that a set of traffic motifs can be taken into account for describing the relevant traffic context used by SoM agent's internal agencies (Table 4.1). These situations can be considered as primitive traffic patterns, which multi-agent traffic simulations of any complexity can be further based on.

The methodology used for this investigation is similar to the one used in the previous section for standard car-following evaluation. In each situation both the vehicle of interest c and its neighbours are SoM driver agents initialised with certain values of personality traits. For each traffic motif, the activation rate of internal agencies for driver c is observed in relation with drivers' innate behavioural propensity (the personality pattern).

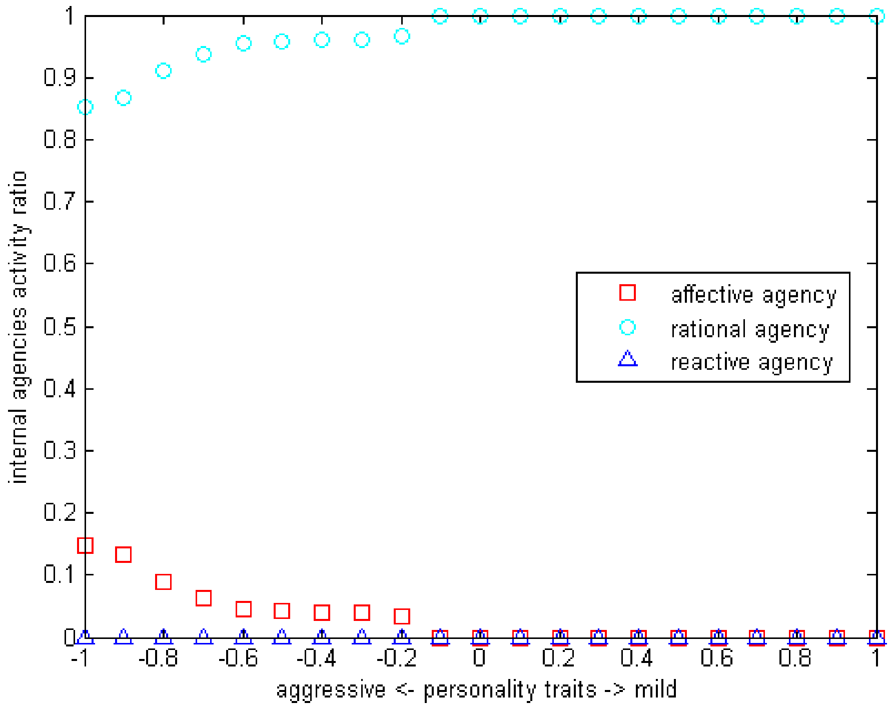


Figure 4.9 Agency activation rate: Motif 1.

Results of the simulations show that indeed the activation rate of each internal agency is consistent with both innate and contextual inputs. Figure 4.9 (motif 1) depicts how when lacking interaction with neighbours the agent behaves close to rationality (optimality) even when its personality traits are at the negative extreme. Despite personality factors participating in emotional update function, there is no contextual addition due to lack of neighbours. Hence, the resultant decision is mostly rational even for the most negative personalities, and as a result, the deviation from optimality is kept to a minimal level.

Figure 4.10 shows the activation patterns for motif number 2 to motif number 9, with increasing amount of interaction: from vehicle of interest and one neighbour to vehicle of interest and all four neighbours. The activation pattern is consistent with observations made for the standard car-following setup. It shows an increase in activation of affective agency as the interaction with neighbours increases (i.e. number of neighbours). An increase in interaction with neighbours translates into more contextual distance inputs contributing to emotional update function in addition to contribution of personality traits. Thus, the activation rate of affective agency decreases slower with positiveness of personality when the amount of interaction with neighbours increases. This equals in term of graphical representation with an enlargement of activation equilibrium (affective-rational) interval. This changes from a biased interval (-0.4;-0.2) for motifs with one neighbour to a symmetrical interval (-0.4;0.4) for motifs with three and four neighbours.

The above interpretation of the results can be counterintuitive, showing that the reason for such behaviour stays in the fact the equations used for simulation are mainly based on distance between neighbours. Hence, whenever there is no neighbour, the equation will definitely go to rational condition at the steady state. Indeed, the equations are mainly based on distances to neighbours, but not entirely. Thus, if there is no neighbour, that does not mean the behaviour goes entirely rational. Drowsiness can appear as a result of no contextual input when the road is free of other drivers, which from a cognitive perspective means there is nothing to raise the alertness and focus levels of the driver. This behaviour is contained by the equations, and hence a behaviour that is different from the intuitive expectation.

Apart from the affective-rational discussion, agent's consistency also results from the activation pattern of reactive agency. Motifs with high interaction with front neighbours (2, 5 and 7) produce higher rates of activation for reactive agency when compared to motifs with high interaction with back neighbours (3, 6 and 8). This is due to the fact the front-free motifs provide either free drive or free adjacent lane. The chances of collision with front neighbour decrease, and so does the need of emergency breaking generated by reactive agency.

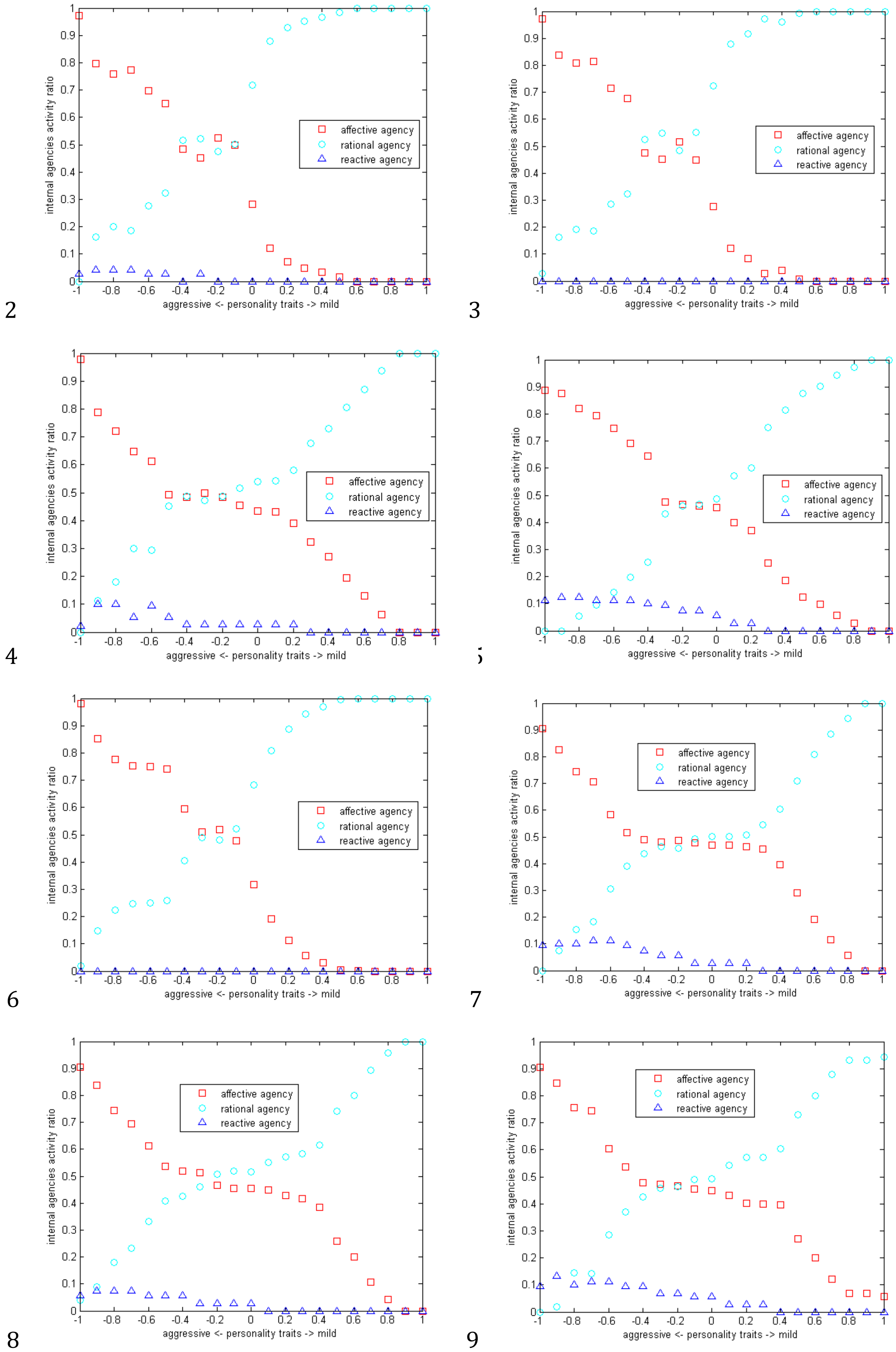
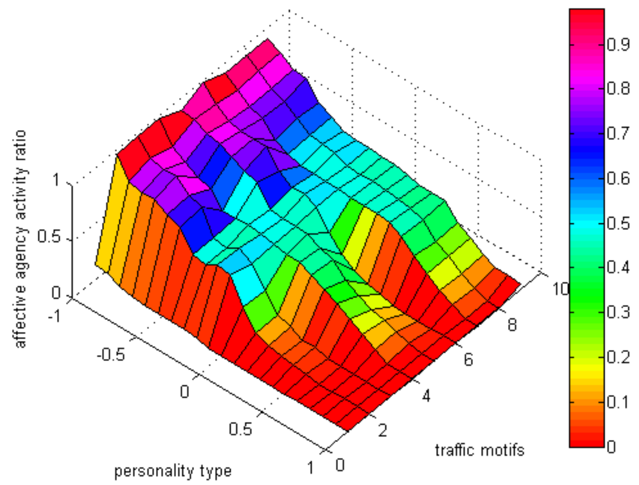
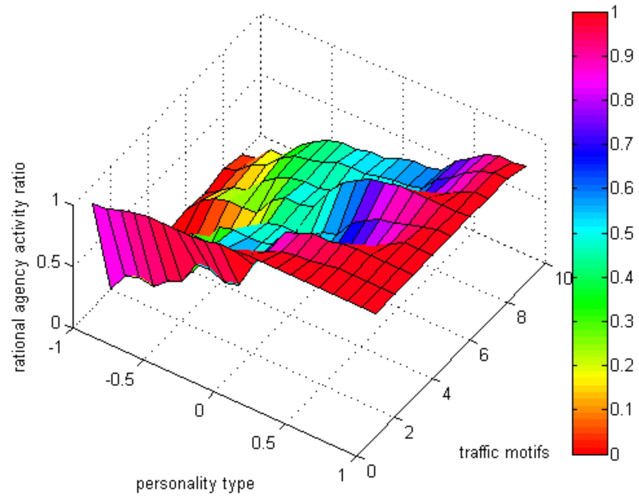


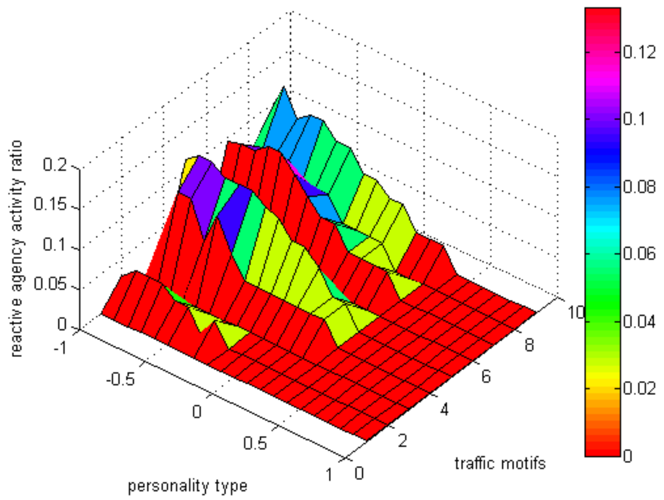
Figure 4.10 Agency activation rate: motifs 2 to 9: □ – affective agency, ○ – rational agency, Δ – reactive agency.



a)



b)



c)

Figure 4.11 SoM driver agent – activation landscape over the traffic motifs and personality types: a) affective agency; b) rational agency; c) reactive agency.

Agency activation landscape. In order to provide a better understanding of the overall agency activation pattern in relation with both personality type and traffic motif, a 3-dimensional activation-personality-motif representation is displayed in Figure 4.11. The figure shows in how each internal agency activates over the "landscape" generated by agents with various personalities in various instantaneous traffic situations. The activation rate of affective agency (a) increases overall towards negative personality and high interaction motifs. It has a global maximum in the high-interaction <> negative-personality corner, while in the same time it decreases at a lower rate towards the high interaction side.

Activation of reactive agency has also an overall increasing pattern towards negative personality and motifs with high interaction, with a global maximum in the high-interaction <> negative-personality corner. In the same time the overall activation landscape shows ridges along motifs with high-forward low-backward interaction, and valleys along those with high-backward low-forward interaction.

4.3. Discussion

This chapter presented a plausible implementation of the non-hierarchical hybrid SoM agent architecture proposed in Chapter 3.

The proposed general architecture is evaluated in a cognitively demanding context, the car driver agent, showing that an SoM driver agent can produce a wide range of driving behaviours in several car-following situations. The SoM agent was tested in a set of car-following situations and demonstrated consistency with traffic expectations from a standard car-following point of view. Also, an investigation of agent's internal agency activation pattern demonstrates the consistency of its structure and internal dynamics.

Arguably, the capability demonstrated by the proposed SoM agent architecture, through its driver agent instantiation, is appropriate for use in more complex road traffic situations. It could be used in multi-agent setups for purposes such as system-level investigation of traffic psychology and behaviour, or assessment of traffic performance and safety measures.

It can be hypothesised that usage of populations of SoM driver agents with various distributions of personality will allow investigation traffic performance for understanding the influence of human users on system performance. These aspects will be treated in detail in this thesis in the next two chapters.

With regard to the research questions of this thesis, this chapter treated partially the research question number 2.

Chapter 5. Investigation of road traffic performance using SoM driver agents in a multi-agent setup

Knowing the influence of driver behaviour on traffic performance is a key issue in transportation. In infrastructure planning, largely used conventional equation-based traffic models (Brackstone & McDonald, 1999) and agent-based models (Panwai & Dia, 2007) emphasise on modelling physical laws (motion) of driver actions under the assumption of rationality. They do not touch the virtually infinite range of non-rational type of actions generated by the variety of human behavioural patterns. Approaches like these may, arguably, not show the real levels of traffic performance. Hence, the resultant system design may be not very useful or even totally unfit with the real needs of the beneficiaries.

Chapter 4 showed how the SoM driver agent, as an instantiation of the more general SoM agent architecture presented in Chapter 3, was able to model a variety of driver behavioural patterns. This fact gives confidence that the approach is usable for a pertinent assessment of the involvement of collective driver behaviour in performance of road traffic networks.

In this chapter the individual SoM driver agent presented is used in a multi-agent setup for investigating the behavioural aspects involved in system-level performance of road traffic networks.

The chapter is organised in four sections. First section presents the general experimental setup. Second section investigates the influence of population size on traffic performance. Third section investigates the influence of behavioural pattern of driver population on traffic performance. Last section concludes on chapter's contribution and draws the conclusions.

5.1. Experimental setup and investigation methodology

The experimental bed-test for all investigations performed in this chapter is based on the approach described in Chapter 4 for evaluation of individual SoM driver agents. A simplified traffic simulator is created following the design principles used in some of the most popular traffic simulators, such as VISSIM (Fotherby, 2002) and AIMSUN (Barcelo, et al., 1998). Road segments are defined through the spatial (geographical) coordinates and vehicle motion uses a continuous representation of the physical space. The vehicles adjust their position at each time-step of the simulation according to instantaneous speed resultant from SoM agent's internal dynamics. Hence, the simulator allows the import and usage of real GPS/GIS maps data in which road segments are defined by geographical waypoints. However, in this chapter only an artificially generated simplified map is used, for being more relevant to the proposed traffic behaviour investigation.

5.1.1. Street map and traffic regulations

The artificially generated map (Figure 5.1) implements a road network consisting of roads with two lanes per direction in the shape of a rectangular grid with 16 (4x4) road junctions modelled as 4-way uncontrolled intersections. Priority, passing and lane usage rules are as in the Australian formulation of the general asymmetric traffic regulations in use worldwide (Kesting, et al., 2007):

- left lane is the default lane. The right lane should only be used for overtaking, or entering road junctions if the direction of movement requests it;
- passing through a road junction follows the "left-hand side" priority rule;
- maximum "legal" speed is 60 km/h (also used below to describe free-flow traffic for calculating some of the traffic performance measures);
- maximum deceleration used for computing the ideal minimum following distance and the ideal minimum gap acceptance for lane-changing is -9 m/s^2 (Kesting, et al., 2007).

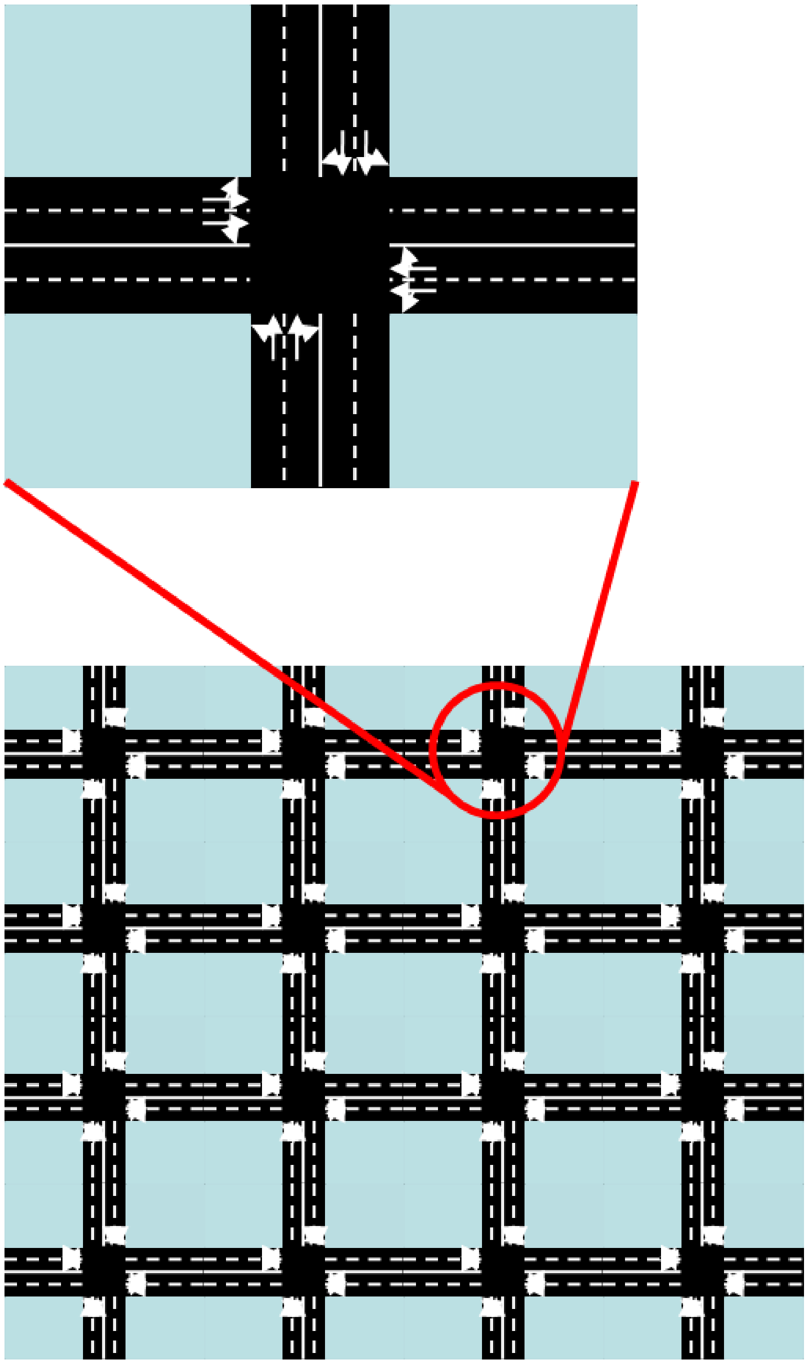


Figure 5.1 A 4x4 grid network with uncontrolled road junctions

5.1.2. Traffic demand, Origin-Destination matrix and routes

Since the street map is a regular network the inhabitants are considered to be uniformly distributed in space, and hence the traffic demand follows the same type of distribution. Thus, the population of drivers used in simulations is spatially assigned uniformly to the existing map and subsequently the Origin-Destination

(OD) matrix is also generated following uniform distribution. Then, for each individual the route from Origin to Destination is calculated using shortest path approach, with random choice where multiple paths of same length exist.

In the next step the behavioural propensities (personality traits) are assigned to individuals in the population of drivers according to a certain statistical distribution in order to investigate the resultant traffic behaviour. The traffic simulator is run for populations with various sizes and statistical behavioural inputs and the appropriate traffic behaviour measures are calculated, as explained in the following sections.

5.1.3. Driver populations

In relation to the existing map and traffic assignment explained above, driver populations can be investigated in two directions: the size and the personality (behavioural pattern) mix.

Size. Population size has an obvious influence on system performance through that it creates either a relaxed or a busy traffic. In this study the artificially generated map is used with populations of sizes multiple of 100 between 100 and 1000 drivers.

Personality mix. Behavioural pattern of an individual driver is implemented, as explained in Chapter 3 and Chapter 4, using personality traits as the expression of innate behavioural propensities. As a reminder, for each SoM driver agent personality was implemented using the BIG FIVE personality model as a five dimensional tuple $P(P_O, P_C, P_E, P_{\boldsymbol{A}}, P_N)$, with $P_i \epsilon (-\infty; \infty)$, where P_i are the personality factors. Computationally, the infinite interval was represented by limiting values of personality factors P_i to a finite real interval $P_i \epsilon [-P_{max}; P_{max}]$, where P_{max}=1. Discussion presented in Chapter 4 showed how different personality traits generate different behavioural patterns for individual agents. If the discussion is extended to more individuals, the resultant population of drivers is expected to show different collective behavioural patterns depending on the mix of individual personalities. These mixes can be generated as simple as multiplying an individual driver agent and obtaining a purely homogeneous population made of individual drivers with identical personalities. Also, in a more realistic view, personalities of individuals can follow certain statistical distributions resulting in populations with various degrees of heterogeneity. Figure 5.2 shows an example of how personality traits of individuals are initialised within the population of drivers, according to a certain statistical distribution. In this figure, a generic normal distribution is presented. The process is similar for normal distribution, or other distributions.

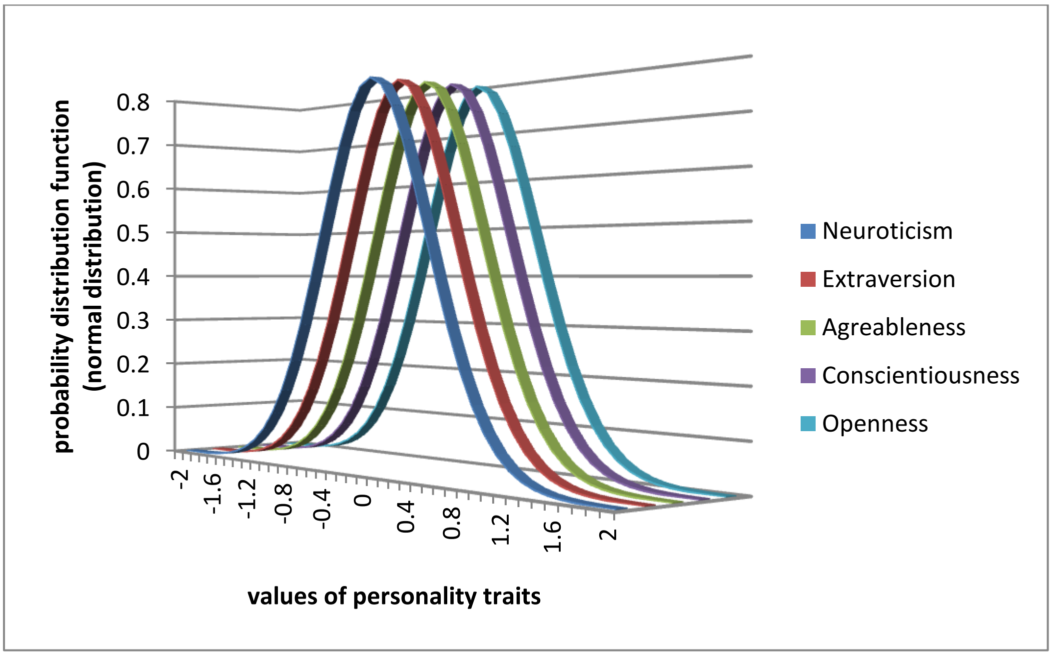


Figure 5.2 Normally distributed values for each personality trait of individuals within a population of drivers

In this study both homogeneous and heterogeneous mixes are treated: homogenous populations of identical driver agents, and heterogeneous populations with uniform and normal distribution of individual personality traits.

5.1.4. Traffic performance measures

Using various populations of drivers such as those described in previous paragraphs, the traffic performance is assessed and discussed using a reduced set of system-level traffic measures. These measures were found by several studies to be of critical importance (Canada, 2006): *average travel time*, *average travel speed*, *travel time index* and *total delay*. They briefly are explained below.

Average Travel Time (t_T)

Average travel time is defined as the average over the whole population of drivers of the individual travel time, where individual travel time is the time for a vehicle to travel between two points (i.e. Origin and Destination in this study).

Average travel speed (v_a)

Average travel speed is defined as the average over the whole population of drivers of the individual travel speed, where individual travel speed is the average speed for a vehicle to travel between two points (i.e. Origin and Destination in this study).

Travel Time Index (TTI)

Travel Time Index is defined as $TTI = t_{Ta}/t_{Ti}$, where t_{Ta} is the actual average travel time of the population, and t_{Ti} is the average travel time in ideal free flow conditions (i.e. constant speed of 60km/h in this study).

Total Delay (D)

Total delay is defined as the total time spent in traffic above the ideal time recorded in free-flow conditions (i.e. constant speed of 60km/h in this study).

5.2. Investigation on influence of the population size

In a first step it is important to investigate how the performance related traffic measures are influenced by the population size. It is also important to establish how this influence is related to the expected behaviour of SoM driver agents in a multi-agent setup.

From a traffic performance perspective it is expected that traffic quality decreases when the number of vehicles on the road increases. This fact is confirmed by simulation results shown in Figure 5.3, where all traffic measures considered in this study deteriorate when population size increases. Additionally, Table 5.1 presents the exact values for the trends presented in Figure 5.3.

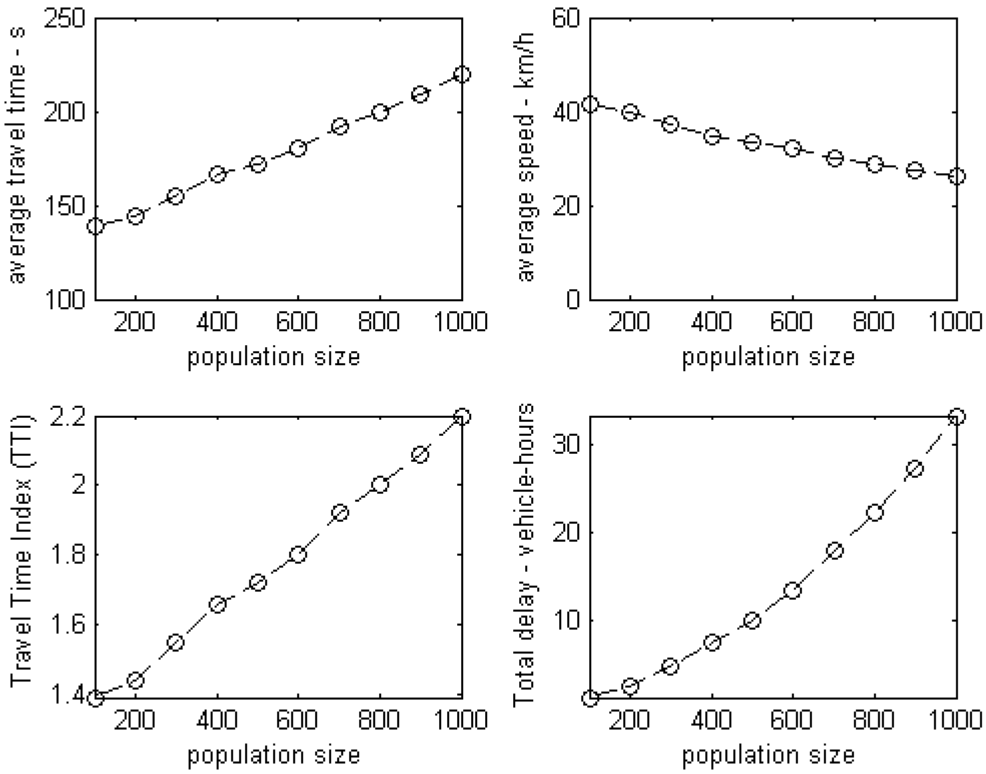


Figure 5.3 Traffic measures for various population sizes: average travel time (t_T), average speed (v_A), Travel Time Index (TTI), Total delay (D).

population size	100	200	300	400	500	600	700	800	900	1000
t_T	139	144	155	166	172	180	192	200	209	220
v_A	41.44	40	37.16	34.7	33.49	32	30	28.8	27.56	26.18
TTI	1.39	1.44	1.55	1.66	1.72	1.80	1.92	2.00	2.09	2.20
D	1.083	2.444	4.583	7.333	10	13.33	17.89	22.22	27.25	33.33

Table 5.1 Traffic measures for various population sizes

However, from the SoM driver agent perspective it is important to understand how decision of individual agents is affected by traffic density and how the dynamics of driver population changes as a result. Results obtained in Chapter 4 showed that for individual SoM driver agents activation rate of internal affective agency increases when agents were situated in busy traffic motifs (for faster reference the traffic motifs are presented again, in Table 5.2). This comes as a result of individual contextual inputs described in the same chapter in Figure 4.4. It can be inferred then, that in the current multi-agent setup deterioration of traffic performance with the population size also appears as a result of increasing pressure on individual drivers when roads become busier. This fact potentially generates an increased amount of affective decisions at the driver population level.

	Motif	Description
1	C	current vehicle
2	C1	current vehicle and neighbour 1
3	C4	current vehicle and neighbour 4
4	C14	current vehicle and neighbours 1 and 4
5	C12	current vehicle and neighbours 1 and 2
6	C34	current vehicle and neighbours 3 and 4
7	C124	current vehicle and neighbours 1, 2 and 4
8	C134	current vehicle and neighbours 1, 2 and 4
9	C1234	current vehicle with all neighbours

Table 5.2 Traffic motifs

Results shown in Figure 5.4 and Figure 5.5 confirm the above statement. Figure 5.5 shows what was expected from a traffic point of view: while the population grows the individual drivers are more and more involved in higher density traffic motifs. For populations of 100 to 300 drivers most of the driver agents are involved in traffic motifs 1, 2, and 3 (zero or one neighbour). As the population size grows the number of drivers involved in high density traffic motifs increases as a result of increased vehicle density. Thus, populations of 700 to 1000 individuals have a high ratio of drivers involved in motif 9 (all four neighbours). Also, for all densities (i.e. all population sizes), the traffic motifs with two neighbours on the same lane (4, 5) are more present than those with neighbours on the adjacent lane (7, 8). This fact that is consistent with the asymmetric traffic regulations considered in the simulations, that is, left lane must be used unless overtaking or very high density. From the SoM driver agent point of view, Figure 5.4 shows how the ratio of affective decision grows with the population size, similar to the ration of involvement in higher density traffic motifs. Consequently, it can be concluded that the activation ratio of individual internal agencies is dependent on the ratio of involvement in traffic motifs. The affective decision increases with traffic density, and so does the reactive decision as a result of riskier behaviour generated by affective decisions, while the rational decision decreases accordingly.

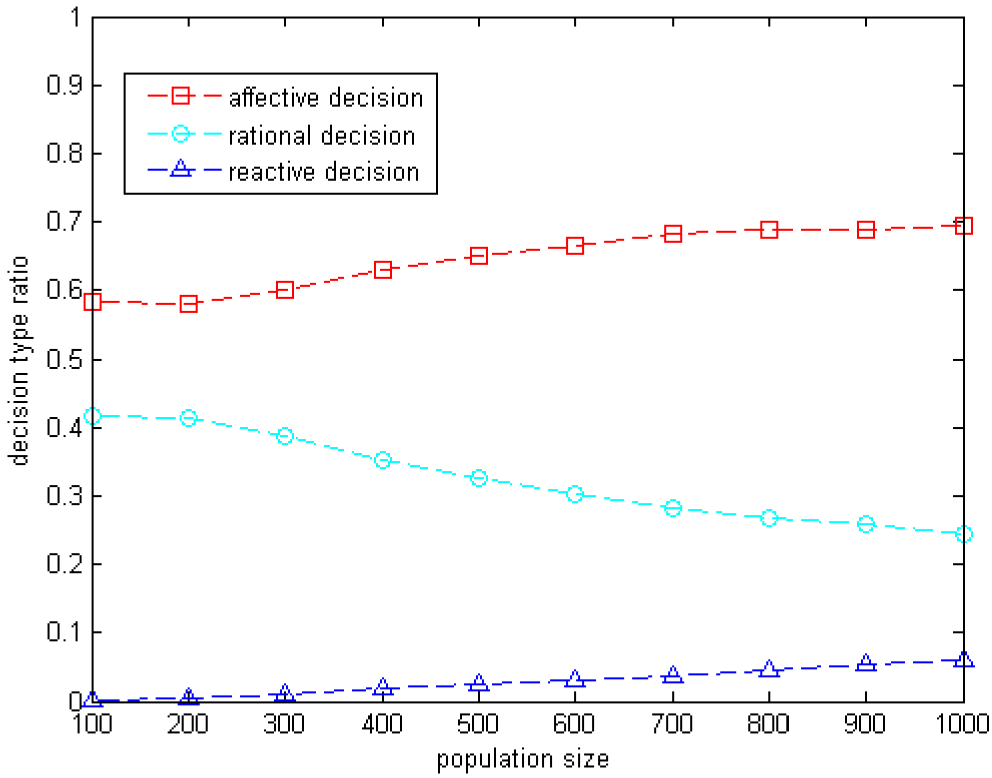


Figure 5.4 Decision type (internal agency activation) ratio <vs> population size: □ – affective decision ratio, ○ – rational decision ratio, Δ – reactive decision ratio.

On one side this conclusion is consistent with results of the individual evaluation of SoM driver agent showing. In the multi-agent setup the SoM driver agents keep the same internal dynamics as in the individual setup, and hence the resultant collective behaviour also follows similar dynamics at population level.

From a different point of view results also show that deterioration of traffic performance is not only a result of the increased vehicle density per se. It is also a result of the interaction between drivers as a consequence of increased pressure induced by the high levels of density.

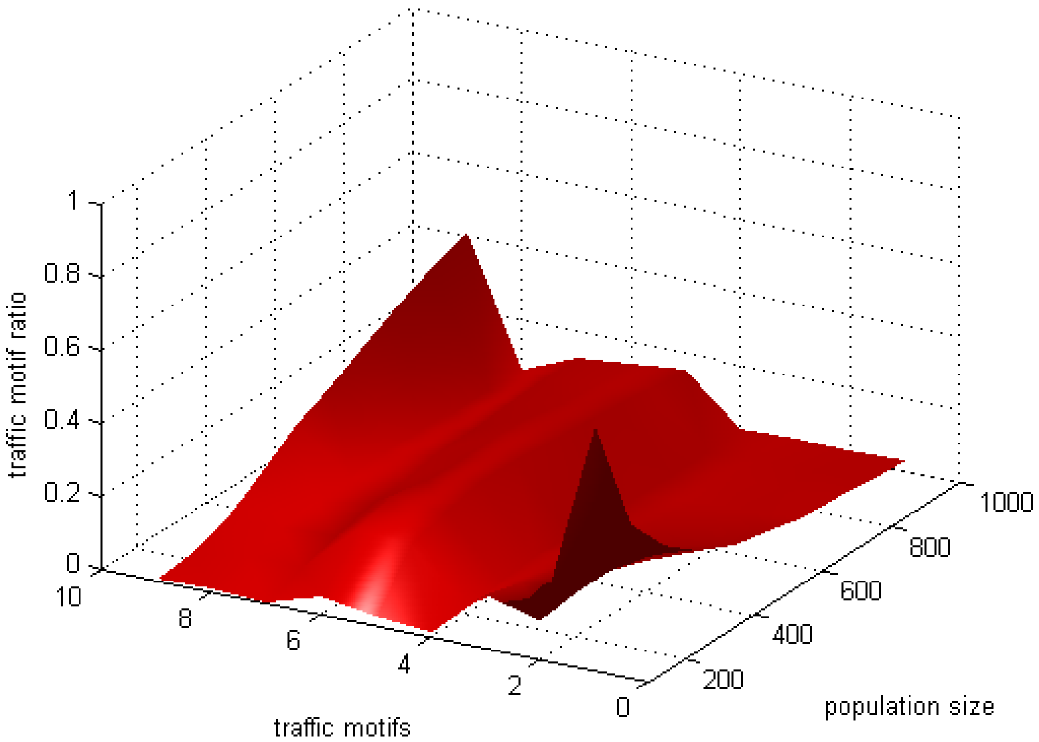


Figure 5.5 Traffic motif ratio per driver population for various population sizes.

However, this aspect needs to be further studied by using more types of population in a high-density traffic context, for seeing how traffic performance alters when the behavioural pattern of the population changes. This investigation is performed in the next section.

5.3. Investigation on influence of the behavioural mix of driver population

In Section 5.2 the investigation was focused on the influence of population size on traffic performance using homogeneous populations of different sizes. These populations consisted of identical SoM driver agents with balanced personality traits (P_i=*0*). However, results suggested that not only the density in itself but also interaction between drivers has an influence on traffic performance. Thus, a more detailed view on the composition of driver populations is needed, in order to understand more clearly how individual behaviour contributes to collective behavioural pattern of driver population.

The investigation is performed for the case of high density, with populations of 1000 SoM driver agents. First, the ideal case of perfectly homogeneous populations is analysed (i.e. drivers with identical personality features). This shows how populations with certain deviation from the balance point of personality, either towards negative of positive side of personality, change the individual behaviour and through this the traffic behaviour. Second, a similar investigation is performed for the case of heterogeneous populations. This should provide an insight on the more realistic scenario in which personalities of individual drivers in a population follow a certain statistical distribution.

5.3.1. Dynamics of homogeneous populations of drivers

Simulations are run for a set of populations with 1000 identical SoM driver agents with individual personality traits P_i sweeping the interval [-1;1] with a step of 0.1 – where -1 represents negative and 1 represents positive personalities.

Results presented in Chapter 4 for SoM driver agents in individual setups shown an increasing activation rates for the internal affective agency as the personality traits deviate towards the negative side of the personality space. Also, a directly proportional variation of non-rational decision rate with the traffic motif density was observed. It was then hypothesised that such a behaviour at the individual level could generate a similar population-level behavioural pattern, determining a deterioration of traffic performance. This hypothesis is confirmed by the current multi-agent setup investigation. Figure 5.6 shows how all traffic measures considered in this study deteriorates as the population personality pattern moves towards negativity.

The pattern of traffic performance deterioration is consistent with variation of collective activation (decision type) ratio for internal agencies of individual SoM driver agents (see page 104). Decision type ratio (shown in Figure 5.7) is calculated based on the total amount of decisions of a certain type (affective, rational, reactive) taken throughout the simulation for all SoM driver agents in the population. Results show how populations situated towards the negative end of personality display an increased affective decision ratio, i.e. higher deviations from rationality. It is confirmed that populations consisting of driver agents with negative personality traits tend to generate an increased overall behavioural pattern that deviates from rationality. This results in deterioration of traffic performance.

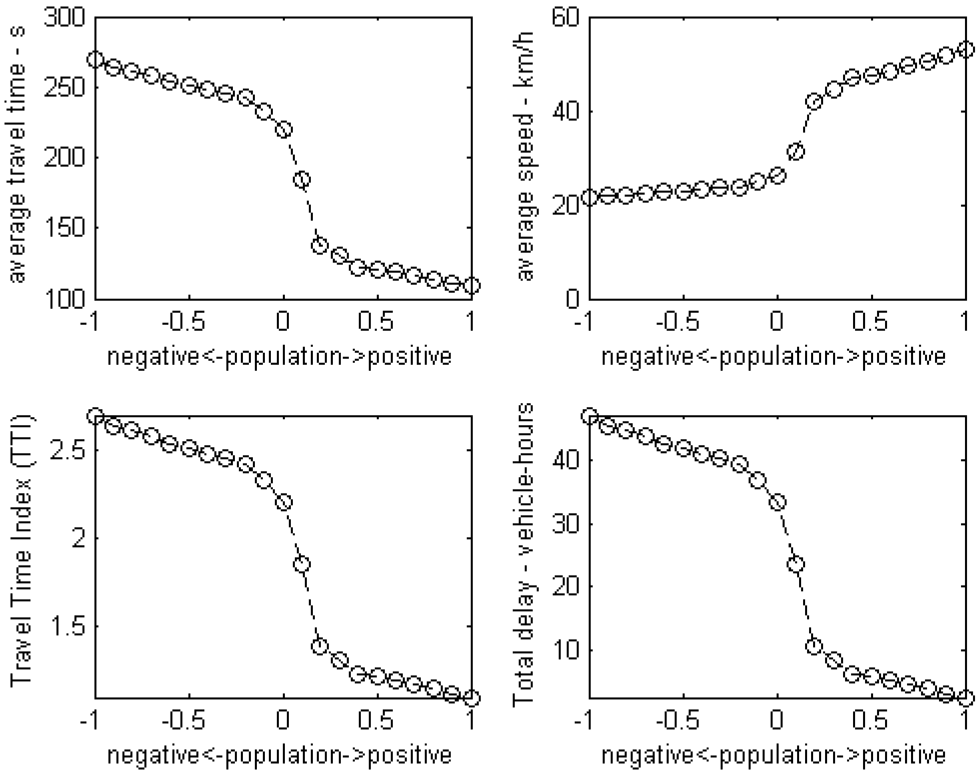


Figure 5.6 Traffic measures for homogeneous populations of 1000 drivers with various behavioural pattern: average travel time (t_T), average speed (v_A), Travel Time Index (TTI), Total delay (D)

In addition to this dependency, Figure 5.8 shows that better traffic performance is obtained in the case of populations of positive individuals even though the amount of traffic motifs of high density (motif 9) is higher. This suggests, in accordance to the findings presented in Chapter 4, that populations of individuals with positive personality can provide high traffic performance even in high-density traffic. As opposite, the populations situated on the negative side generate low traffic performance despite the traffic density being lower.

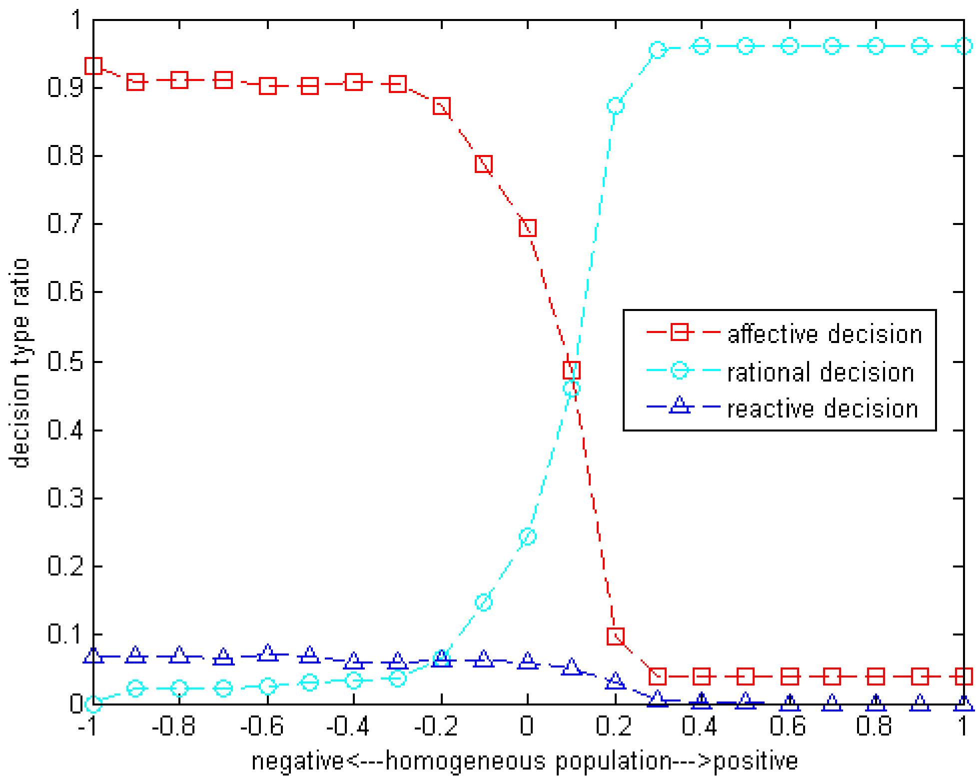


Figure 5.7 Decision type per homogeneous populations of 1000 drivers with various behavioural patterns: □ – affective decision ratio, ○ – rational decision ratio, Δ – reactive decision ratio.

Overall, the investigation on homogeneous populations confirms that traffic performance decreases when density increases not only because of the density in itself. There is also the influence of drivers decisions when facing dense situations (traffic motifs). Analysis showed that in the same density conditions homogeneous populations of drivers with positive personality traits allow better traffic outcomes than those consisting of negative drivers.

Yet, the fact of having a homogeneous population of drivers, be it situated towards positive or negative side of the personality spectrum, is in itself an ideal situation. Indeed, this serves for getting a convincing proof that differences in drivers' individual behavioural propensities generate differences in the overall traffic performance of the resultant population drivers. However, no population in real life has identical individuals. Hence, a more realistic investigation is needed in order to create a complete picture of the influence of collective behaviour on traffic performance. This analysis is provided in the next section, which treats heterogeneous populations of drivers.

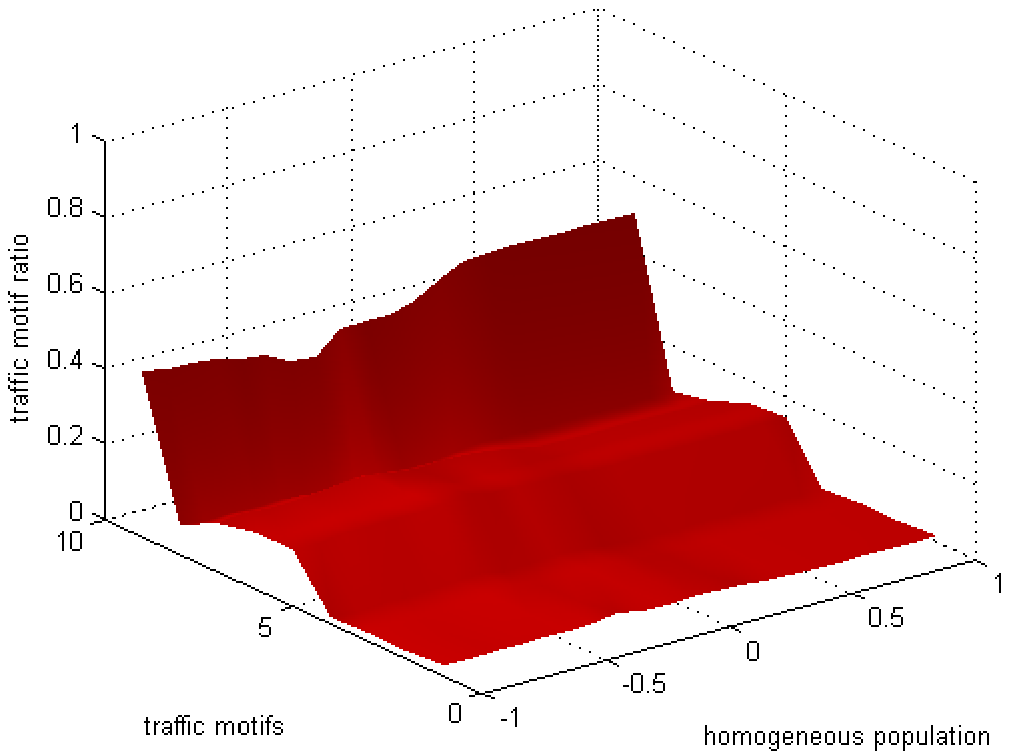


Figure 5.8 Traffic motif ratio per homogeneous populations of 1000 drivers with various behavioural patterns.

5.3.2. Dynamics of heterogeneous populations of drivers

In Subsection 5.3.1 the investigation focused on ideal homogeneous populations in order to show that driver populations with different personality pattern have different behaviour in traffic conditions. However, real populations show various degrees of heterogeneity of personality features depending especially on geographical factors, but not only.

This generate certain patterns in the statistical distribution of personality traits – an aspect which was intensely studied by theory of personality and individual differences (David P. Schmitt, Allik, McCrae, & Benet-Martínez, 2007). Allik and McCrae found that people living in the same country or region show similar or identical means whereas those geographically separated or historically isolated from each-other have less similar means of personality traits (Allik & McCrae, 2004). Standard deviation was also found to dependant on geographical aspects (R. R. McCrae, 2002). McCrae found that Asian and African nations had lower standard deviations from the established means when compared to European and American populations.

From a traffic behaviour perspective, it becomes obvious that an accurate investigation on a real geographical area must take into account the personality distribution of the local population of drivers. In this chapter the artificially generated map is used without a specific geographical localisation. The investigation is performed using driver populations with various degrees of heterogeneity (standard deviation) of personality traits in order to observe the impact of personality distribution on both traffic performance and dynamics of agent population.

Thus, the study is performed for populations of 1000 drivers, in which the majority of individuals has a balanced personality and only a reduced amount of them is situated towards the extremes of personality space. From a statistical distribution point of view it is assumed that personalities of drivers in populations follow normal distributions with μ=0 and σ>0. From a heterogeneity point of view variation of σ from 0 towards ∞ equals with populations ranging from purely homogeneous when σ=0 to purely heterogeneous (uniform distribution) when $\sigma \rightarrow \infty$.

Given that the personality representation is computationally limited to the real interval [-1;1], the investigation is performed for values of σ between zero and one – $\sigma \in (0; 1]$. It is considered that σ=0 represents the balanced homogeneous population investigated earlier in this chapter in Section 5.2, and any value σ>1 can approximate an uniform distribution.

From a traffic performance point of view results depicted in Figure 5.9 show that all traffic measures deteriorate strongly as the heterogeneity, represented by σ, increases. However, it should be noted that for a population of 1000 drivers, the traffic is at high density. Hence, the deterioration of average speed in this case is not significant, since in the most favourable case the average speed is already low.

Additionally, Table 5.3 presents the exact values for the trends presented in Figure 5.9. This finding is important as it shows that dynamics of a heterogeneous population actually jeopardise the fluency of traffic as the heterogeneity increases. Or, in other words, the more diverse are the interacting drivers the more impediment they create for traffic flow resulting in an overall inefficient collective behavioural pattern.

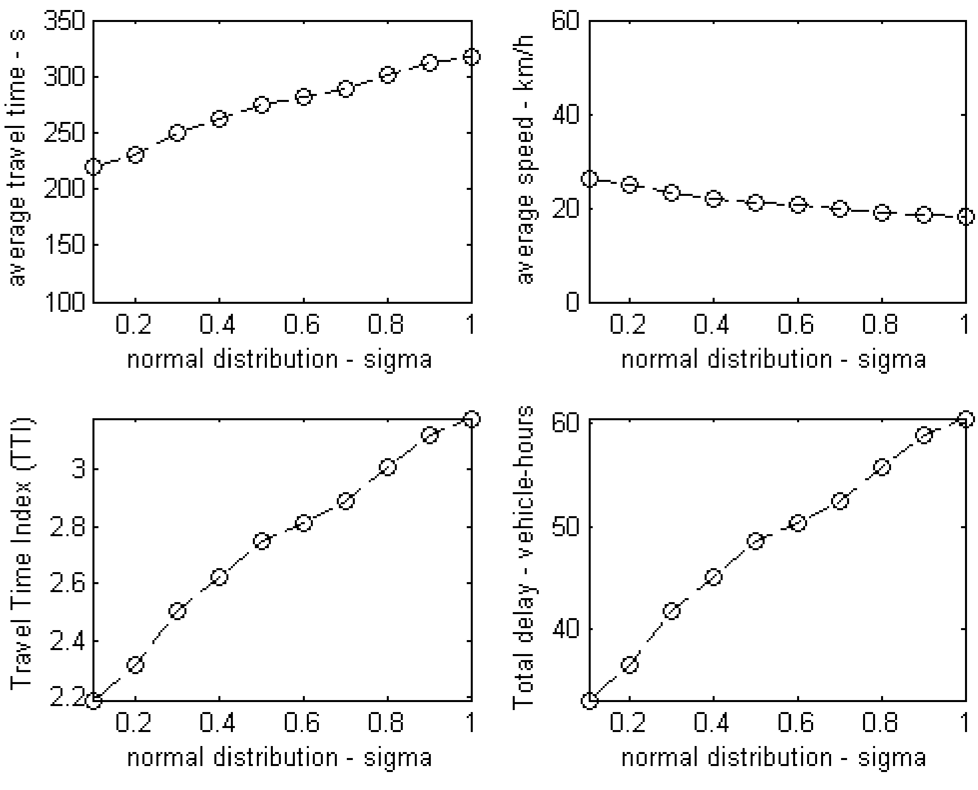


Figure 5.9 Traffic measures for normally distributed populations of 1000 drivers – μ=0, $\sigma \in (0; 1]$: average travel time (t_T), average speed (v_A), Travel Time Index (TTI), Total delay (D).

σ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1**
t_T	219	231	250	262	275	281	289	301	312	318
v_A	26.3	24.94	23.04	21.98	20.95	20.5	19.93	19.14	18.46	18.11
TTI	2.19	2.31	2.50	2.62	2.75	2.81	2.89	3.01	3.12	31.8
D	33.06	36.39	41.67	45	48.61	50.28	52.5	55.83	58.89	60.56

Table 5.3 Traffic measures for normally distributed populations with various σ

This aspect is also supported by data recorded in Figure 5.10, with the ratio of affective decision per population growing on behalf of the rational decision as the heterogeneity (σ) increases. The traffic motif ratios per population (Figure 5.11) also show that for similar involvement in dense traffic motifs populations with lower heterogeneity cope better. These populations generate an overall collective behavioural pattern that allows higher traffic performance.

From a different point of view, the above results also show that deterioration of traffic performance is significantly more important than in the previously discussed case of homogeneous populations. It can be seen that the travel times recorded for heterogeneous populations are in general higher than those recorded for homogeneous populations.

The worst case of homogeneous population – the one corresponding to extreme negative personalities – generates an average travel time t_T=275 seconds (Figure 5.6). In comparison, the worse scenario for normal distributed populations has an average travel time t_T=318 seconds (Figure 5.9). It can be seen from the two cases that the common point from traffic performance point of view corresponds to populations with P_i=-0.5 (middle of negative personality space) on the homogeneous side, and to populations with σ=0.3 on the heterogeneous side.

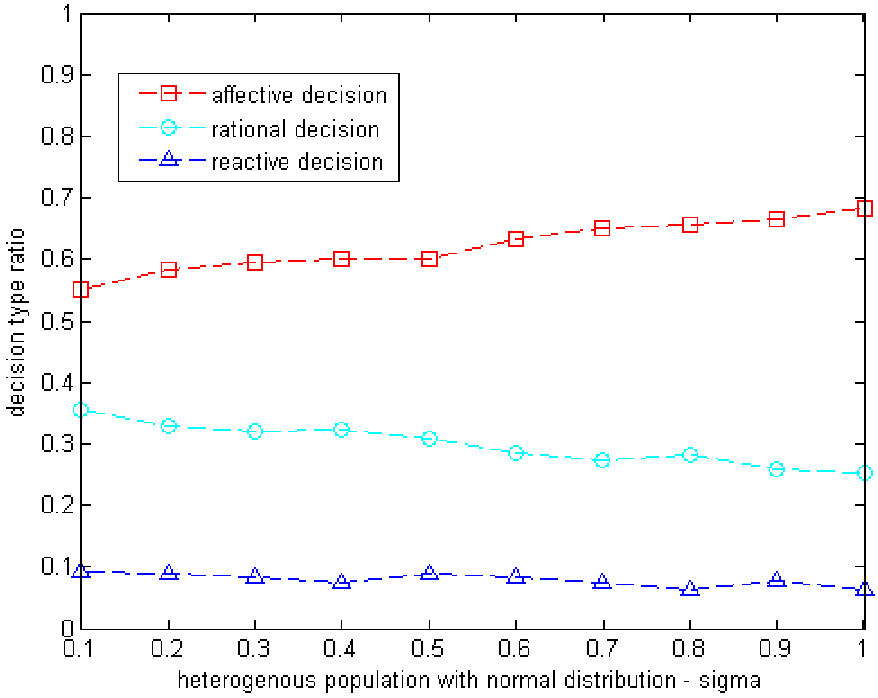


Figure 5.10 Decision type per normally distributed populations of 1000 drivers, μ=0 and $\sigma \in (0;1]$: ○ – affective decision ratio, ○ – rational decision ratio, Δ – reactive decision ratio.

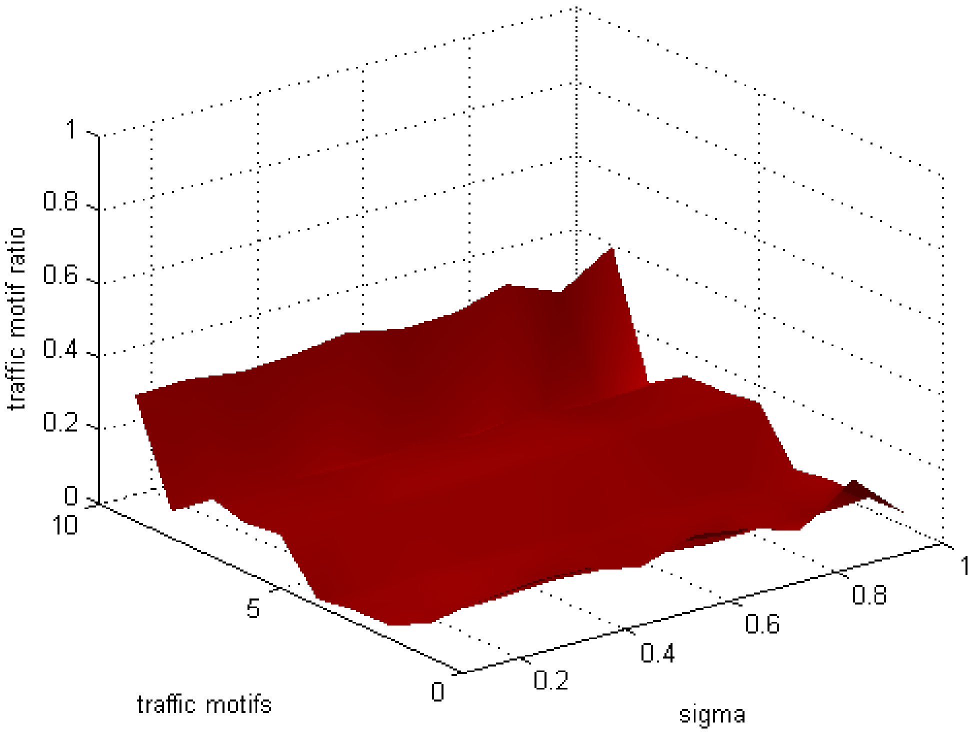


Figure 5.11 Traffic motif ratio per normally distributed populations of 1000 drivers, μ=0 and $\sigma \in (0;1]$.

5.4. Discussion

In this chapter the SoM driver agent, presented and evaluated in Chapter 4, was used in a multi-agent setup in an artificially generated traffic environment. The purpose was to demonstrate the importance of individual behaviour for traffic performance.

First, the influence of the size of driver populations on traffic performance was investigated using populations of identical drivers with balanced personalities. An immediate and obvious conclusion was that larger populations of drivers generate increased vehicle density that deteriorates the overall traffic performance. However, investigation of population dynamics showed that the behavioural pattern is also important for the overall traffic performance, not only the number of drivers.

Second, this finding was further explored using ideal homogeneous populations situated at various distance in the personality space from the balance point (P_i=0) on both negative and positive side. The investigation confirmed that traffic performance is significantly altered when population personality pattern deviate from the balance point towards the extremes of the behavioural/personality space. It was shown that traffic performance decreases as the populations deviate towards the negative side.

Third, the investigation was extended to realistic heterogeneous populations of drivers in which personalities of individual drivers are not identical but they rather follow certain statistical distributions. It was shown that traffic performance decreases significantly when the heterogeneity increases. The overall deterioration of traffic performance for heterogeneous populations was significantly more important than for the homogeneous ones. This demonstrated that interaction of drivers with different personalities leads to lower traffic quality than the interaction between identical drivers.

On one side, from a traffic perspective, the insights gained from this chapter are as follows:

- predictions of conventional non-behavioural traffic models, such as equation-based car-following models or agent-based models which assume rationality, may be significantly in error. This could have important impact on the outcomes of traffic performance assessment, and consequently on the resultant design strategies.

- use of behaviour enabled artificial populations of driver agents could significantly improve planning, assessment and design outcomes. This could have broad implication in both day-to-day urban activity approaches and artificial life autonomous systems approaches on transportation.

From a different point of view, this chapter also demonstrated the usability of the SoM agent architecture proposed in this thesis – through its implementation as a SoM driver agent – in a multi-agent setup. The investigation performed on SoM driver populations with various personality mixes showed that resultant collective behaviour was plausible. It was consistent with possible real-life traffic expectations, as well as with the internal dynamics of the individual SoM driver agents discussed in Chapter 4.

With regard to the research questions of this thesis, this chapter treated partially, and completed the answer to research question number 2.

Chapter 6. Investigation of transport systems resilience using the SoM driver agent: case study on Melbourne

In Chapter 5 investigation performed on an artificially generated road traffic network demonstrated that the collective behaviour of car drivers has an important influence on the resultant traffic performance. An analysis of this aspect was possible by using in a multi-agent setup the SoM driver agent described in Chapter 4, as an instantiation of the general SoM agent architecture proposed in Chapter 3. To this point the thesis showed the usability of the proposed general agent architecture and demonstrated its use in a cognitively demanding environment, as a driver agent in a road traffic behaviour context. However, investigations performed in Chapter 4 and Chapter 5 used artificially generated environments for both individual and multi-agent settings. They emphasised on both internal dynamics of the SoM agents and resultant traffic performance outcome. The next question arising in relation with the SoM agent is whether its driver agent instantiation can be used in real world contexts for investigating more complex transportation issues. One of this issues is the behavioural aspect involved in resilience of road transport systems of real geographical areas (regions) or cities.

In this chapter the SoM driver agent is used in a multi-agent setup in a case study, in order to assess the transportation resilience of city of Melbourne. Here the resilience of a transportation system is seen as an interplay between transport infrastructure and behaviour of population of drivers. This dual nature of a transportation system comes from the fact transportation, as a critical infrastructure, is different from other critical infrastructures, such as electricity or water networks, through that the human element carries a much heavier importance. Humans are themselves parts of the transported commodities and influence through their behaviour the resultant traffic performance. In the electricity or water networks the "commodities" only obey to physical laws and they flow as they are planned to. On the contrary, in transportation humans account for causes but also for effects and they are in the same time planners, active participants and beneficiary. They can be either protective or harmful and thus risk factors affecting the system. On the other hand the system can be itself a risk factor for humans. That is because the surrounding environment creates mental states which affect the decision-making. The resultant behaviour reflects on the further utilisation of the system (and thus on the system itself), generating a complex chain of causes and effects appears.

For this reason research in transportation, with respect to concepts such as performance, risk or resilience, has been always under the sign of duality between the technological and behavioural aspects. Arguably, the transportation system can be virtually seen as operating at these two main levels: technological (infrastructure) and behavioural (Figure 6.1). However, in real life they are so strongly melded and interwoven that making a clear distinction between them, be it just for categorising purposes, is extremely difficult.

Thus, countless attempts to model the transportation systems, predict their evolution, improve their performance or assess the risk involved by their operation have been made over time. Despite the increased interest and efforts, none of the approaches managed to capture the systems in their whole, to produce accurate representations and to generate truly realistic estimations of their evolution. None of the approaches can stand as complete tool, usable by all stakeholders involved in any way in system operation. Arguably, it is the emphasis on the technological aspects and the lack of addressing in detail the influence of the human element which limits the accuracy of the proposed models.

This chapter brings together both technological and behavioural approaches by placing face-to-face resilience measures treating infrastructure and resilience measures treating the behaviour of driver population. It emphasises on the role of humans in the system, suggesting that human behaviour introduces risk in the transportation landscape in at least the same extent as the infrastructure failures do. For this reason treating both aspects as part of planning or performance/resilience assessment processes is highly desirable and extremely necessary. This could eliminate the existing shortcomings in the research on improving the performance and mitigating the risk involved in operation of transportation systems.

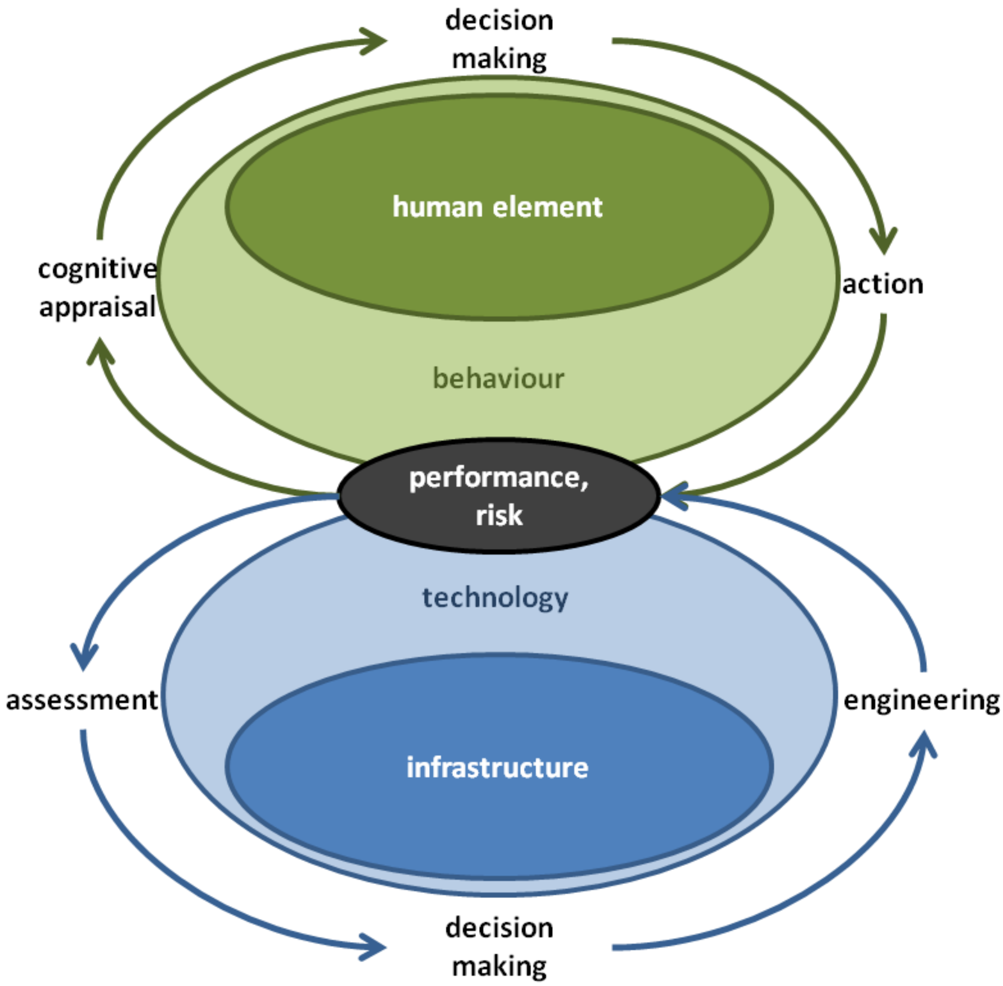


Figure 6.1 Dual nature of the research in transportation

The chapter is organised in six sections. First section provides an overview on the research in transportation resilience as an infrastructure issue. It briefly presents the most important achievements in terms of resilience metrics and design approaches for building resilient transport systems, emphasising on their drawbacks. Second section presents the experimental setup, showing how Melbourne transportation data is extracted from local and global sources and modelled into computationally suitable data structures which support the proposed investigation. Third section provides an in-depth systemic investigation of the infrastructure layer of transportation systems, considering as a case study the city of Melbourne. Various resilience metrics are calculated with regard to loss of network connectivity and cost of re-establishing the affected waypoints and paths in the event of various physical or service failures. These metrics are then used to assess Melbourne transportation system from an infrastructure perspective. Fourth section completes the assessment performed in the second section by switching the focus towards the behavioural domain. The behavioural investigation is based on the SoM driver agent implemented and evaluated in the earlier chapters of the thesis. It brings into attention various traffic metrics and their variation generated by changes in the collective driver behaviour. Through this, the current chapter also becomes an evaluation of the SoM driver agent usability in a city-scale setup. Fifth section brings together the findings from third and fourth sections. It introduces a resilience index, which includes both infrastructure and behavioural measures of system performance. The purpose is to create a comprehensive resilience assessment framework for road transport systems. Last section, the sixth, draws the conclusions on the main findings and contribution of this chapter.

6.1. Overview on transportation resilience

Resilience has been defined in many ways over the years, but none of the definitions was comprehensive enough to capture the variety of existing systems in a single view. However, resilience analysis and assessment are nevertheless essential parts of any pertinent process of risk assessment and management in the nowadays systems. Researchers worked over the years with a variety of concepts, such as vulnerability, reliability, resilience, robustness, efficiency, sustainability and many others. Their purpose was to analyse, understand, model and improve various systems, be they manmade or natural. All these concepts are irrefutably related, and gathered under the wider umbrella of risk. The most important of the concepts will be briefly discussed in the following paragraphs.

Jenelius (Jenelius, 2007) tries to explain the similarities and differences between three main concepts “vulnerability”, “robustness” and “resilience” which are considered essential parts of any risk analysis. First, he highlights that vulnerability analysis, as part of risk analysis considers intentional attacks or threats and investigates their impact on the serviceability of the system itself. As a broader concept, risk analysis may include these aspects but rather considers "random" unintentional events, hazard that is. It investigates their impact on elements that can be external to the analysed system (e.g. surrounding environment, competitors, customers etc.). Yet, it should be noted that Jenelius sees the risk as the "*combination of an event with negative consequences for human life, health or environment and the probability of this event*". For this reason, in his study risk and vulnerability are different but also have similarities. They are somewhat completing each-other in the process of assessing the risk, where risk is seen as a sum of negative elements. Robustness and resilience are presented in the same study (Jenelius, 2007) as properties of the system itself, excluding the external environment. Robustness is seen as "*the ability of a system to retain its functions when is under strains*", while resilience is "*the ability to recover to a normal state after having been disturbed*". He considers that vulnerability implies a lack of robustness and resilience, while the vice versa is not necessarily true. Hence, he suggests that robustness and resilience are "the converse" of the vulnerability.

Villagrán De León (Villagran, 2006) underlines that resilience if defined by some researchers in the same way with robustness. He says that some believe resilience is the ability of a system to resist the impact of a given event. This actually the robustness according to Jenelius. Others believe that resilience is system's ability to absorb and cope with an event, so that the term of resilience took roots in the literature and somehow included the meaning of robustness. Hence, it became generally accepted in the literature the investigation of the relationship between vulnerability and resilience. De León also concludes that in the light of this relationship vulnerability becomes the reciprocal of resilience. Thus, if a given system is highly resilient then it has a low vulnerability. Folke and colleagues also describe resilience as the "*flip side'* of vulnerability, but he then suggests that resilience is a determinant of vulnerability (Folke, Carpenter, et al., 2002; Folke, Colding, & Berkes, 2002). In a different vision the Resilient Organisations (McManus, Seville, Brunsdon, & Vargo, 2007) claims that vulnerability is "*a* *measure of the criticality, preparedness, and susceptibility of the components of an organisational system*". As a result, vulnerability is seen as a component of the resilience.

Resilience is thus an evolving concept that tends to replace all the other concepts used in risk management and planning by incorporating them. Hence, the investigation of system vulnerability, as part of resilience analysis and assessment, magnifies the problems that need attention. Then, the remedies for these problems become the strategies that the system needs to adopt to increase its level of resilience.

In the field of transportation, resilience is mainly treated in terms of reliability. Most of the literature in the field acknowledges that resilience can be analysed using two main classes of methods: those related to connectivity reliability and those related to travel time reliability. Connectivity reliability (Bell & Iida, 1997) is calculated mainly using a binary model – functional and non-functional – for nodes or links. These are part of particular paths defined by pairs of source-destination (S-D) nodes. The probability that path S-D is functional, as a measure of reliability, is computed according to specific status of the established nodes or links. Time travel reliability (Clark & Watling, 2005) is also based on probabilistic study of the nodes and links, but more from the perspective of their usage and the effect on the travel time associated with the paths they belong to.

Other classes of methods have been introduced by different researchers. Chen and colleagues (Chen, Yang, Lo, & Tang, 2002) defined capacity reliability as "*the probability that the network can accommodate a certain traffic demand at a required service level*". According to this definition, reliability can be seen as a complex interplay between the quantity of flow and the quality of services. Behavioural reliability (Watling, 2002; Yin & Ieda, 2001) takes into account the effect of the drivers' behaviour on the general performance of the network. This promotes the idea that reliability is the result of a game played by all the participants in the operation of a transportation system. Another class of methods attempts to analyse the potential reliability (Berdica, 2002; Jenelius, Petersen, & Mattsson, 2006). Here, the aim is to assess the risk implied by network operation through the identification of weak points, vulnerabilities, planning flaws and their effects at system level. These methods emphasise on vulnerabilities as key issues in the analysis of resilience.

Another approach for analysing resilience in the field of transportation arises from the type of the event that generates the need of analysis. All methods discussed above take into account events which are small scale and likely to happen in the day-to-day system operation. These events can be minor accidents, road maintenance or failures in traffic signalling. Other methods tend to take into account large-scale disasters that trigger major damage at system level, such as major earthquakes, floods, volcanic activity, etc. Sakakibara (Sakakibara H., 2004) examined the resilience of transportation networks in terms of "robustness" against catastrophic disasters. The investigation was performed as a case study on region in Japan, which was subject to major damage produced by earthquakes. Sakakibara introduced the "topological index", a special system-level measure for topological reliability which is appropriate for large scale damage characterisation.

In Australia, Evans and colleagues (Evans, Burke, & Dodson, 2007) consider that most of the planning and reliability assessment of Australian transportation systems is based on older methodologies. They highlight that Melbourne Integrated Transport Model (MITM), Brisbane Strategic Transport Model (BSTM) and Strategic Transport Model of Sydney (STM) are all mainly based on the Four Stage Travel Demand Model (FSM) (Mathew & Rao, 2007; McNally, 2000) or variations of it. They agree that using such models was and still is inevitable, but they argue that the increased complexity involved by the resultant management systems makes them expensive. They also criticise the extremely high specialisation needed for their operation, which makes them accessible to a reduced set of highly trained staff/boards/companies. Their doubts about the future of these models are somewhat confirmed by the authorities and the organisations involved in transportation systems' planning. Victorian authorities (VPPIA, 2008) and other organisations (CMTT, 2007) signal that new challenges regarding fast population growth, and increasing people's demands request more efficient and cost effective frameworks. These should provide simpler and more elastic evaluation models for the future.

6.2. Experimental setup

In order to perform the proposed resilience assessment of ground transportation system in Melbourne, real data extracted from GPS navigation maps of Melbourne metropolitan area are used.

In all studies so far, the definition of a network in a transportation network has not been challenged. In the traditional view, nodes in the network are viewed as stations or crossroads, while links are the rails or streets that connect them. Analysing networks modelled in this manner can be suitable for some applications, but is certainly questionable in terms of resilience. That is because from an infrastructure perspective most of the incidents tend to take place on the way, between stations or crossroads. Also, from a traffic behaviour point of view, various traffic measures reflect the quality of traffic on specific segments of the roads rather than exactly at the road junctions. As such, the tendency to represent road-junctions (stations, for rail networks) as nodes and road segments between junctions as links can make it difficult for analysing the true resilience of a network. The traditional junction (station)-based representation of transportation networks may lead to misperception in where the vulnerability really exists. For example, in the case of a train network, the majority of the network is tracks exposed outside the stations. From one perspective, a failure on a track does not depend on where this failure occurs. From another perspective, it does since this track may cross different areas with different population sizes. A failure occurring in a point that is close to a tram track or a bus route would have different impact from one occurring in an isolated location. Thus, an alternate way of defining the transport networks is needed.

The GPS data come with an extra feature. A node in a transportation network is defined by a change in the network rather than the traditional definition of being a station, or a road junction. Hence, the graphs used in this investigation are built as follows. Any node in the network is a waypoint defined by waypoint ID and position (latitude and longitude), while links are directed pairs of waypoints. A waypoint is not necessarily a station in the case of rail network and not necessarily a crossroad in the case of road network. Thus, the resultant graphs for rail and road networks do not follow the layout of station-based models currently used for planning and operation (e.g. Melbourne public transport system – www.metlinkmelbourne.com.au). It is expected that a network defined as in the GPS-based approach can represent the system in a more detailed manner, allowing a more effective investigation of the resilience of transportation systems.

The real data necessary for network analysis is obtained from O*pen Street Map* system, by accessing Open Street Map website (www.openstreetmap.org) and exporting the area of interest (Melbourne area) in XML format. The tram, bus and train traffic networks are extracted from the XML database mapped over the physical infrastructure. The physical infrastructure represents the real geographic map, which pictures the spatial position of all waypoints that are part of the transportation system. The traffic networks for each of the transportation means (e.g. tram, bus, train) represent the path of all routes, correlated with the geographical position of the waypoints. Then, the database for traffic is connected through the position of the nodes to the database for physical infrastructure. Further, the official website of Melbourne public transport system is used for extracting the average frequency of vehicles for each route of each transport mean with regard to peak and off-peak hours. Also, data about spatial distribution of population is taken from the Victorian Government population report (VPPIA, 2008). In the end all data sets are connected to the existing physical layout (i.e. the geographical map of Melbourne area), creating in a single database a GIS hypercube of Melbourne transportation system.

On top of the infrastructure representation, road traffic is also added to the above GIS structure. In order to perform the traffic behaviour assessment, populations of car drivers implemented using SoM driver agents are considered in a similar manner as in Chapter 5. Populations of SoM agent drivers with various degrees of heterogeneity are used for computing traffic/resilience measures over the express roads of Melbourne, i.e. streets with speed limits of 100, 80, 70 and 60 km/h. Personality traits within the driver population are considered to be normally distributed with μ=0 and various values of σ, with $\sigma \in (0; 1]$. Distribution of personality traits for Melbourne drivers is considered normal with μ=0 and σ=0.7 based on previous findings in personality theory (Allik & McCrae, 2004; R. R.

McCrae, 2002; R. R. McCrae & A. Terracciano, 2005; David P. Schmitt, et al., 2007). In these studies authors found that personality traits for Australian population has a standard deviation situated towards the higher end (higher heterogeneity) of the standard deviation interval from a number of 56 countries around the world.

The methodology described above allows a system-level risk identification process which paints the resilience picture for Melbourne City ground transport system. The approach can be generalised to any geographical area covered by a GPS navigation map, being a cost-effective, systemic and structured approach to quantify and manage system performance and resilience.

6.3. Resilience assessment at infrastructure level

Investigation of infrastructure resilience involves two major aspects. First, the infrastructure is treated at physical layer for measuring the resilience given by intrinsic topological properties of the physical road network, such as connectivity, robustness to waypoint failure etc. Then, resilience of the services running on the physical layer is assessed by considering several operational situations. For both aspects a variety of metrics and scenarios are used and the investigation is performed for train, tram and street networks. The interdependency and interaction of these networks, as well as their relation to spatial distribution of the population, are then discussed in detail.

6.3.1. Resilience of physical layer

For the investigation of physical layer, data are extracted from the XML database as follows. Rail network is divided in train and tram networks, and road network is divided in four networks according to the existing speed-limits: 60, 70, 80 and 100 km/h. On the railway side the train network consists of 4675 nodes and the tram network has 3632. On the road side, the street networks consist of 2448, 1525, 2083, and 4212 for streets with speed limit of 60, 70, 80 and 100 km/h respectively.

6.3.1.1. Structural measures

First set of measures captures the network structure in itself. The most significant structural measures reported in the literature are the average degree, the betweenness centrality and the clustering coefficient (Newman, 2008). These measures can highlight possible connectivity reliability flaws induced by the structural properties of a network, without considering operational issues such as traffic or failures.

Average degree. The average degree in a network is a simple and intuitive local measure which gives an idea about the local connectivity of nodes. It is calculated by adding the number of nodes each node is connected to and by dividing this sum by the total number of nodes. In essence, it represents the average number of connections a node has.

Betweenness centrality. Betweenness also shows the importance of a specific node but takes into account the global influence of other nodes. Betweenness of a node i is defined as the fraction of shortest paths between other nodes that pass through node i. It emphasises that a node may be involved in more or less paths between randomly selected nodes in the network. A node with betweenness much higher than the average (if exists) could act like a bottleneck, and it is likely to be a structural flaw especially if alternative routes are not provided.

Clustering coefficient. The measure which intrinsically contains the amount of alternative routes available over the network is the clustering coefficient, which is also calculate in order to complete the view on structural performance. The clustering coefficient simply indicates how much the transitivity relationship holds in a network, i.e. if x is connected to y and y is connected to z, then how often it is found that x is also connected to z.

Data displayed in Table 6.1 show that for all the analysed networks, the structural performance is relatively low. The average degree shows that all networks have relatively low connectivity and, additionally, they contain nodes with betweenness roughly ten times higher than the average. This suggests the existence of bottlenecks within their structures, fact confirmed by the extremely low values of the clustering coefficient. This places the analysed networks in the category of tree-like or even pure tree structures. Taken together, these aspects highlight one possible (and very important) structural flaw: lack of alternative routes.

network type	**degree**		**betweenness**		**clustering coefficient**
	max	**average**	**max**	**average**	
Train	5	2.0273	0.3873	0.0308	0.0007
Tram	4	2.0297	0.3043	0.0389	0.0027
Street 60	5	1.991	0.0579	0.005	0
Street 70	4	1.9685	0.0206	0.0016	0
Street 80	4	1.9759	0.1449	0.0121	0.0008
Street 100	4	1.9829	0.0353	0.0046	0.0011

Table 6.1 Structural measures – degree, betweenness and clustering coefficient

Yet, it should not be neglected that the tram network seems to have the highest structural performance from the whole set of analysed networks. It can be seen that it has the highest average degree and also the highest clustering coefficient. In the same time though, the tram network also exhibits the highest average betweenness among the analysed networks. It can be argued that a high average betweenness signals a high probability for the existence of bottlenecks, fact that should raise some questions about the structural performance of the tram network.

Also, it should be noted that betweenness centrality is based on shortest paths. Consequently, a high betweenness indicates the concentration of a high amount of "shortest" paths through the nodes of interest, without excluding the existence of alternate longer paths. For this reason, the clustering coefficient was used as a complementary tool for assessing the existence of alternate paths. Clustering coefficient for tram network is significantly higher when compared to the rest of the networks, indicating that alternate paths for the potential bottlenecks are indeed available. In conclusion, the tram network can be considered as having the highest structural performance.

6.3.1.2. Measures for robustness to waypoint failures

The second set of measures is based on the concept of node removal, described in (Albert, Jeong, & Barabasi, 2000) and (Leu & Namatame, 2007). Starting from (Leu & Namatame, 2007) two measures are used: Topological Integrity and Distance Gap.

Topological Integrity (T_i). In topological integrity, the number of non-overlapping sub-graphs in a network is computed. The procedure starts with the initial network (graph), from which nodes are gradually removed. After a node is removed, this graph may become disconnected and the removal of that node may split the network into two or more components. For computing the T_i, failures are simulated for each node of the analysed network and then the effect of node removal on network fragmentation is observed, by counting the number of components or sub-graphs (Albert, et al., 2000; Barabási, Albert, & Jeong, 1999). It is likely that some of the nodes fail and leave the network unaffected, while others cause the network to break in several stand-alone disconnected pieces. Then it is calculated the probability that removal of a node k breaks the network in n pieces and results are transferred into a probability density function representing the topological integrity measure. The cumulative probability distributions of these PDFs are also used in the investigation.

Distance Gap (D_g). The case in which failure of a node disconnects the network by breaking it into pieces raises the question of damage cost associated with the failure. In this case, the distance gap created by the removal of a node is calculated. Assume there are three nodes x, y, and z, with y connected to both x and z. When y is removed, the distance between x and z is calculated (remember, each of these nodes has a longitude and latitude associated with it). This distance is used as a proxy for estimating the cost of re-establishing network connectivity. In order to do this assessment, first step is to consider the distance gap produced by the removal of each node. Then the probability that a removal of a node k generates a distance gap d is calculated, and the corresponding estimated probability density function of the generated distribution is visualised.

Figure 6.2 depicts the probability density function drawn from estimating the topological integrity measure. All analysed graphs, except the tram network, show a fairly similar behaviour, with slight differences among them. The probability density function shows a low probability that failure of an arbitrary node leaves the network undamaged. It also shows high probability that failure of an arbitrary node breaks the network in two disconnected sub-graphs. The probability that removal of a node disconnects the graph in more than two separate fragments is very low. For tram network, the plot shows a much higher resilience compared to the rest of the networks.

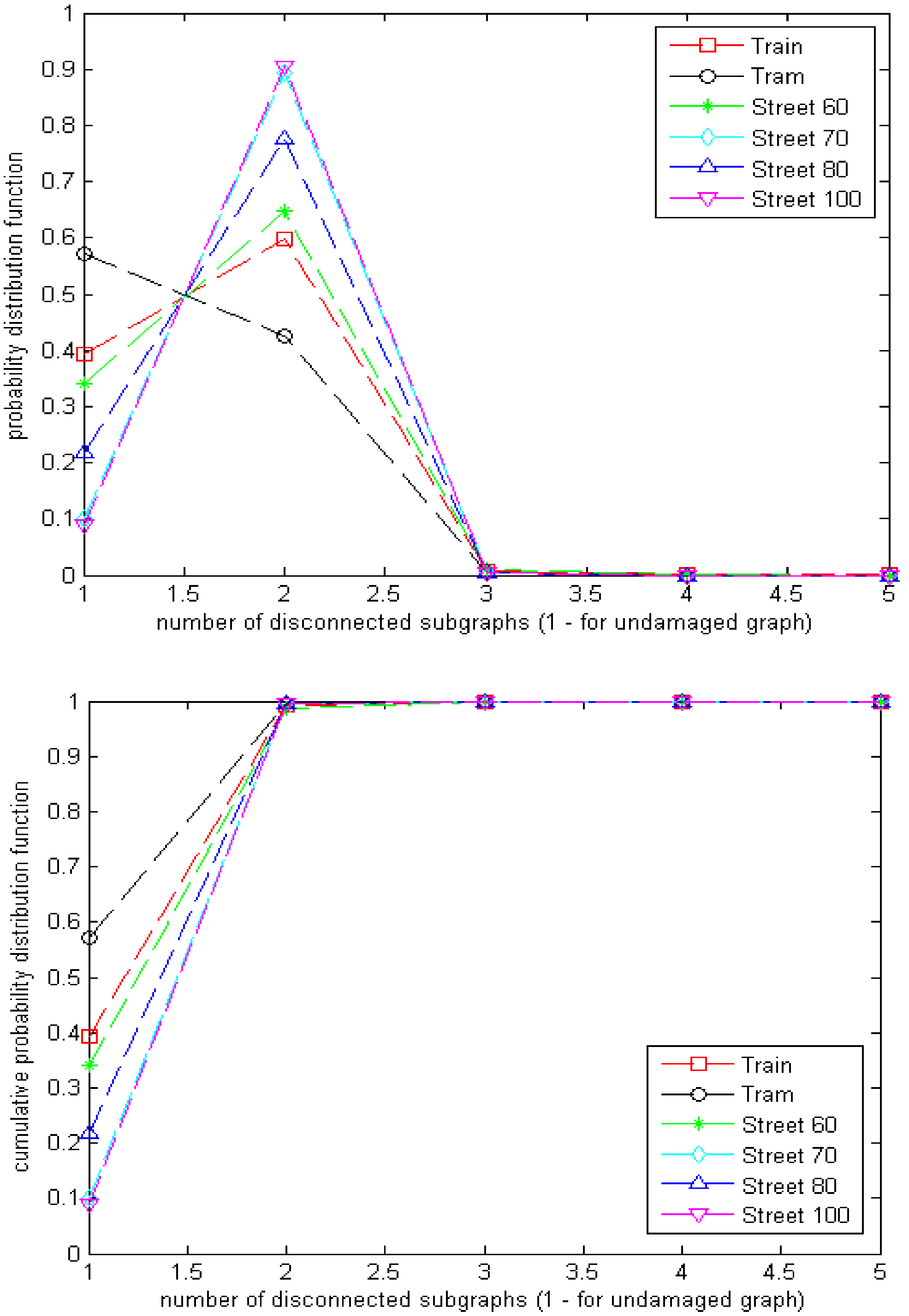


Figure 6.2 Topological integrity – probability density function (top); cumulative distribution (bottom)

Probability that the graph remains undamaged is much higher than the rest of the networks, while in the case of some damage being produced, the probability is significantly lower. It can be also seen that the maximum T_i for tram network is 2 while for all other graphs the maximum T_i reached 5.

The concept of first degree stochastic dominance (Goodwin & Wright, 2009) is pertinent in this situation. The cumulative distribution for Tram network fully dominates (better in all points) the other distribution. Based this the Tram network can be considered as the most robust (resilient).

Topological integrity is a good enough albeit intrinsic measure of resilience. As it is, the structure of the graph offers information about possible vulnerabilities of the network. However, the information is only related to lack of redundancy, or to presence of highly connected nodes that bond different pieces of the network acting like bottlenecks.

Real networks, though, are located in space and their resilience implicitly depends on the distances among different waypoints. In the case of disintegration of the network as a result of arbitrary failures of nodes, distance gaps between disconnected pieces can be calculated, in order to estimate the amount of damage. The distance gap gives a good idea about the cost implied by the recovery of a gap, if cost is considered as a function of the length of the damaged path (e.g. rail to be maintained or replaced, road to be consolidated or rebuilt).

In the case of calculating the resilience (robustness) from distance gap point of view, results can be seen in Figure 6.3. Distances are rounded to 100 metres. It can be seen that the maximum distance gap produced by network disintegration is around 900-1000 metres. All analysed graphs show a fairly similar behaviour, with slight differences among them. As the probability density function does not provide a clear understanding about which network is more resilient, the cumulative probability distribution function is again used. However, because the cumulative curves intersected with each other, the second order stochastic dominance is used for comparing networks' resilience. Second order stochastic dominance requires the calculation of the areas in each intersection between two curves to calculate dominance.

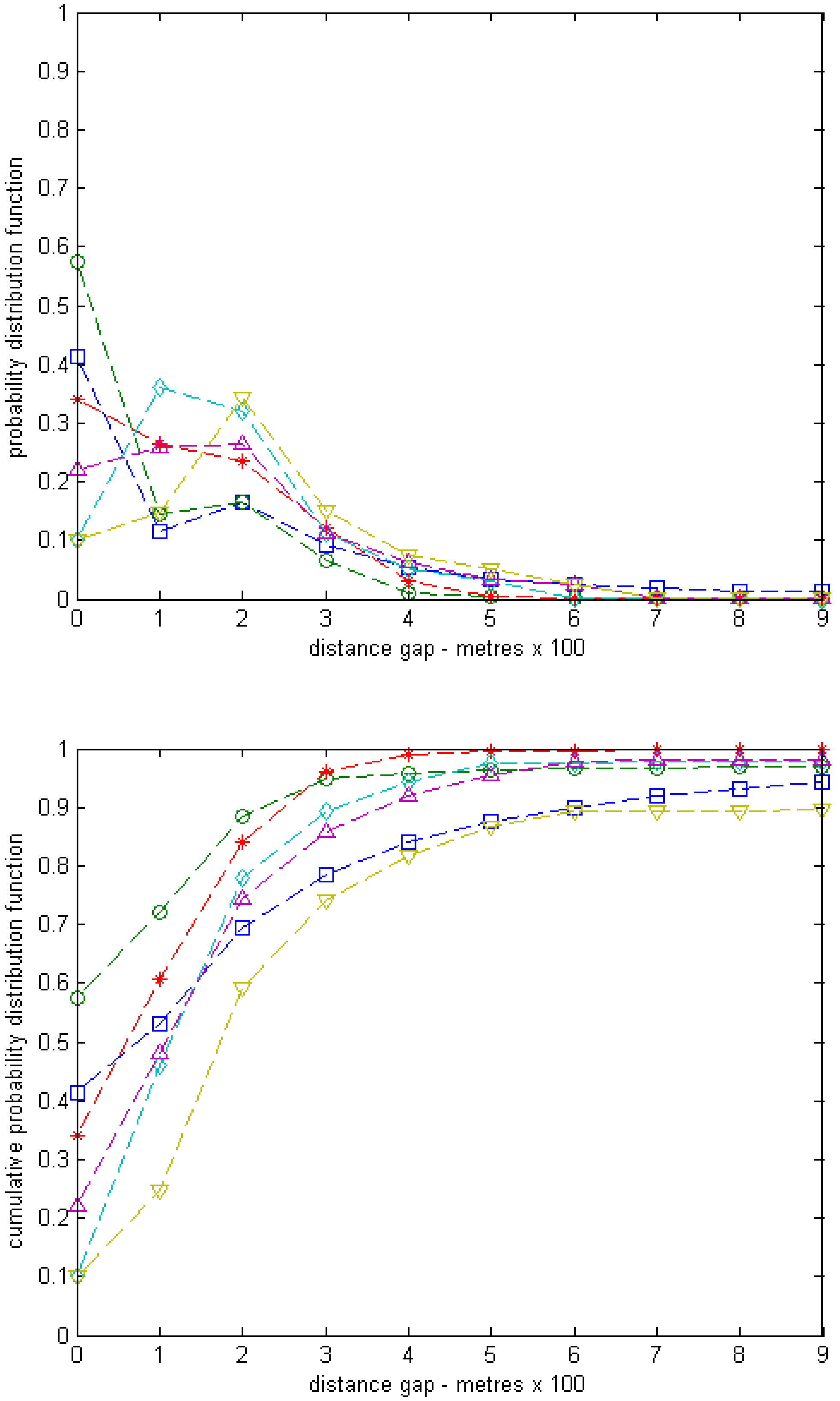


Figure 6.3 Distance gap – probability density function (top); cumulative distribution (bottom)

The resultant order of resilience based on the distance gap measurement shows that the most resilient network is again the tram network followed by the train and streets networks.

6.3.1.3. Measures for spatial distribution of risk

It should be noted again that both the structural measures and the failure related ones are usual tools for network analysis and they account only for the physical level of the infrastructure. They provide valuable information about potential structural or operational flaws. However, they do not offer much information about the overall spatial distribution of vulnerabilities over the system.

Hence, a third class of measures focuses on the spatial distribution of risk. Both topological integrity and distance gaps are mapped spatially to their original location in the longitude and latitude coordinate. Contours of length 5 km are drawn around the city centre as concentric “zones” in order to represent the geographical centrality (distance to city centre). Topological integrity and distance gaps are calculated and visualised versus geographical centrality in order to assess the spatial distribution of potential damage. Nodes with similar topological integrity and distance gaps in each suburb are grouped together. Then, the population size of that suburb is used as a proxy for the possible impact of node failures on local population.

First, the visual information regarding spatial distribution of failures is presented in Figure 6.4. The figure depicts in latitude-longitude coordinates all nodes from all networks coloured based on topological integrity. The image presents the resilience information as a virtual layer over the concentric zones (i.e. geographical centrality) and administrative suburbs. The first inner circle is centred in Melbourne city centre and has a radius of 5 km. Each subsequent circle represents a radius of 5 km away from the previous circle. The blue dots represent nodes with T_i=1, the red dots correspond to T_i=2, and the black dots represent nodes with T_i>2. The figure demonstrates that vulnerabilities decline for zones further away from city centre, confirming an expected outcome, since most transportation activities occur closer to the city centre.

However, Figure 6.4 only demonstrates the information visually. In order to objectively estimate the resilience level, the visual information needs to be transformed into measurable quantities. In order to do that two measures are introduced for investigating the impact of node failures on concentric zones, as follows.

Distribution of T_i=2 over geographical centrality. This measure shows the amount of nodes with topological integrity of 2 in each geographical centrality zone. Figure 6.5 shows the cumulative distribution of vulnerable nodes with T_i=2 in each network over the zones.

Once more, it is clear that the vulnerability of these networks is concentrated in or closer to the city centre. A 15 km radius around the city contains almost 100% of the tram vulnerable nodes, while a radius of 30 km contains close to 90% of all vulnerable nodes including – surprisingly – streets with 100 km/h speed limits.

Impact of T_i=2 on population. The last measure is the impact of a vulnerable node with topological integrity of 2 as measured by the size of local population (Figure 6.6). For each network nodes with topological integrity of 2 are counted and separated based on the suburbs they are part of on the geographical map. Then the average population size affected by these failures is computed and graphically represented. Thus, this measure represents a proxy for the impact of vulnerable nodes on local populations.

Figure 6.6 shows a high impact of the road networks with speed limit of 60 and 100 km/h on the population, from the population size point of view. This could be explained through that (Figure 6.7) the roads with speed limit of 100 km/h are the main radial roads connecting the inner to outer parts of the city. They connect the highly populated peripheral residential areas to the city centre area, and hence, the highest impact on the population. In a similar manner, the roads with speed limit of 60 km/h appear to be a mesh of main roads mapped over the metropolitan area. They connect the metropolitan residential areas, arguably, with high impact on the served population.

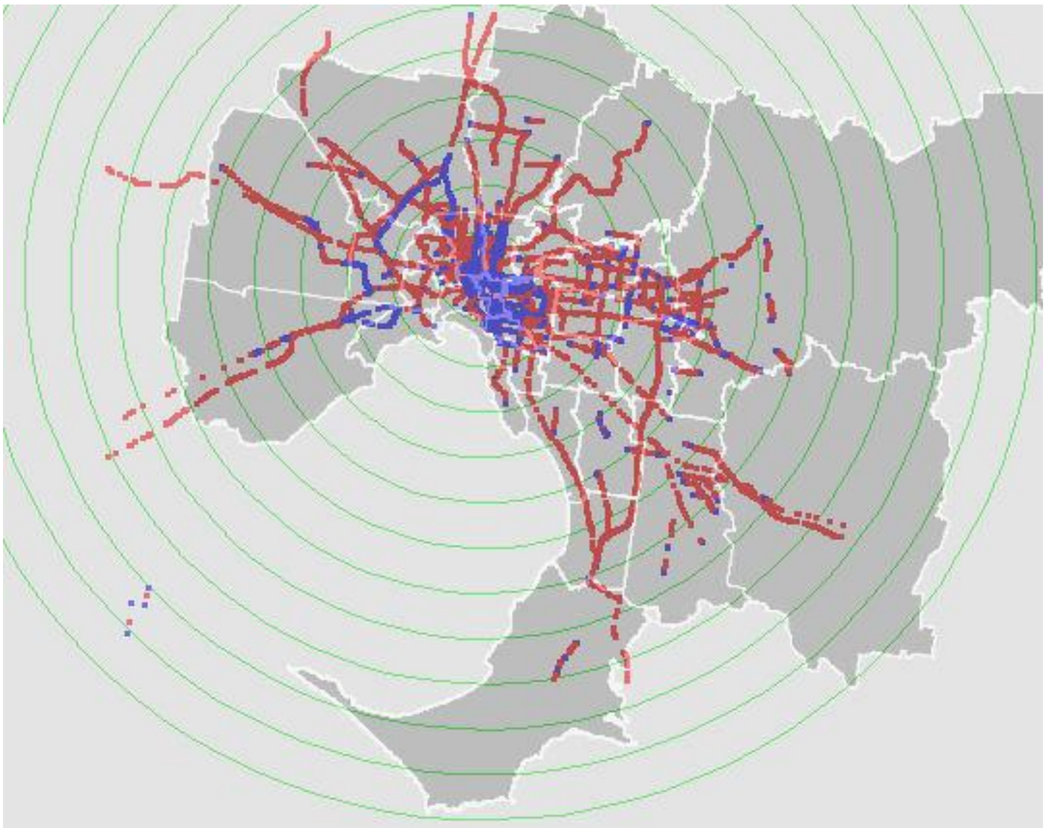


Figure 6.4 Spatial distribution of topological integrity: ■ – T_i=1, ■ – T_i=2, ■ – T_i>2.

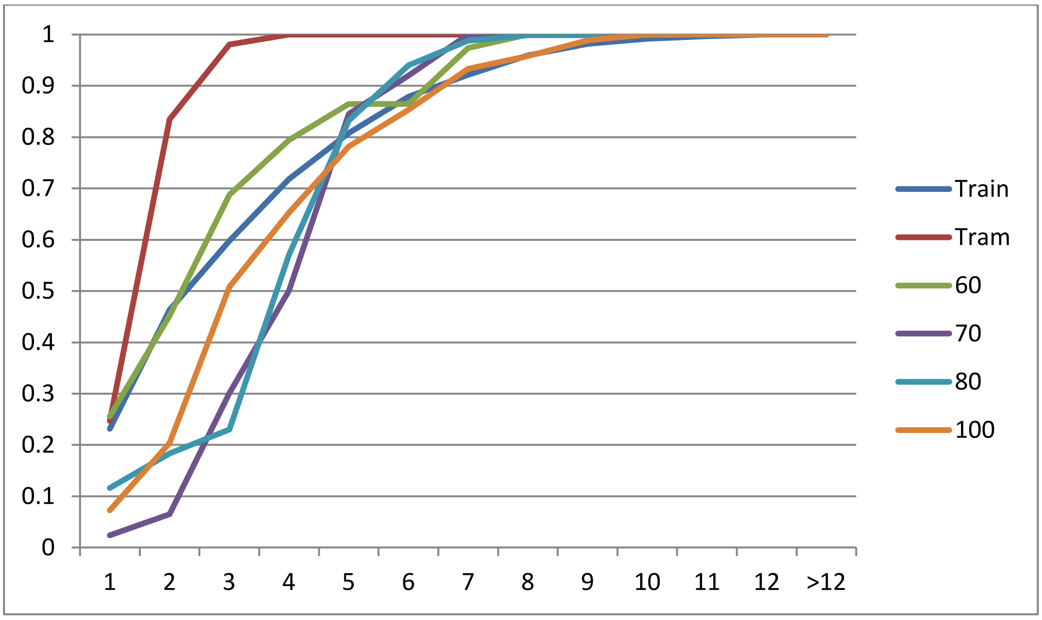


Figure 6.5 Cumulative distribution of vulnerable nodes with topological integrity of 2 over geographical centrality zones

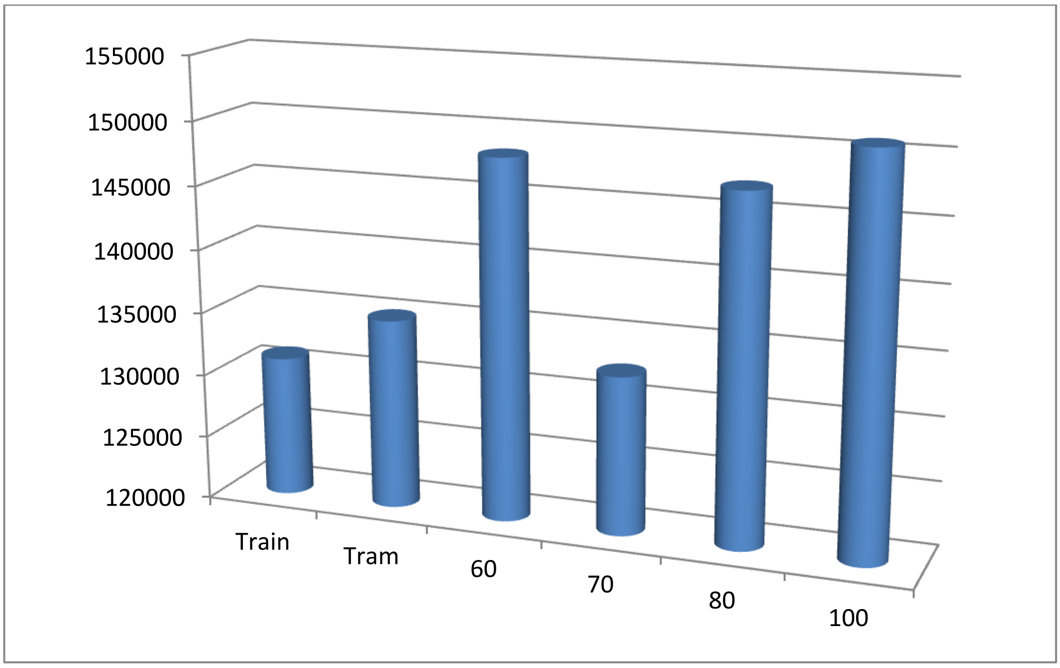


Figure 6.6 Average population size in areas with nodes having topological integrity of 2

Figure 6.7 Street networks with speed limits of: ■ – 60km/h, ■ – 70km/h, ■ – 80km/h, ■ – 100km/h, ■ – all rail network.

6.3.2. Service level resilience

For this investigation the public transport networks (trams, buses, and trains), routes and timetables are extracted from the available data sources. This is done in a similar manner as for the physical networks, as explained in Section 6.2. For each network, the paths for all routes are embedded into the graph-based GIS database, adding to the physical infrastructure.

Then, three scenarios are considered for assessing the damage generated by arbitrary failures of vehicles (trams, buses and trains) in both time and space. In these scenarios it is assumed that a vehicle fails in some point on its route. This forces the passengers to either wait for the next vehicle or walk to the closest point from where they can take a different transportation mean and wait there for the next vehicle. As an example, if a tram carriage fails, then the tracks will be unusable for all the route numbers that pass through the failure point. Waiting for another vehicle belonging to those routes is of no use. In this case, passengers must walk to the closest station of a different transport mean.

6.3.2.1. Scenario 1. Unconstrained distance gap - U_{dg}

This scenario has a minimal equivalent in real world. However, it provides nevertheless a valuable insight for a preliminary assessment. It assumes failures of vehicles in arbitrary points of their routes. Then the minimum distance that passengers of the failed vehicle need to walk to the next transportation mean is computed. In practice this translates into finding the closest waypoint belonging to a route of any of the tram, bus, or train networks.

For this case, the investigation does not take into account any constraints related to real utilisation of the transportation means. Thus, it is assumed that if a vehicle fails in a waypoint on its route, passengers can be picked up by any other transportation mean in the closest point to the failure point. Distance gap is calculated for all waypoints belonging to all networks and then the system level interruption measure is displayed as the probability density function (PFD) of the overall failures. The PDF shows the probability that failure of a vehicle in a point of its route generates for the passengers a distance gap of ‘x’ metres.

Figure 6.8 depicts the probability density function drawn from estimating the distance gap for scenario. It can be seen that in most cases the passengers will find an alternate transportation mean within 50 metres from the failure point. In the worst situation the alternate transportation mean will be within 450 metres. This means that the tram, bus and train routes are close enough to each other to provide alternate transportation within the visual range or, worst case, within one or few blocks distance.

However, this conclusion comes as a result of an ideal case where no constraints are applied to existing networks, whereas in real situations a set of constraints must be taken into account as in the next scenario.

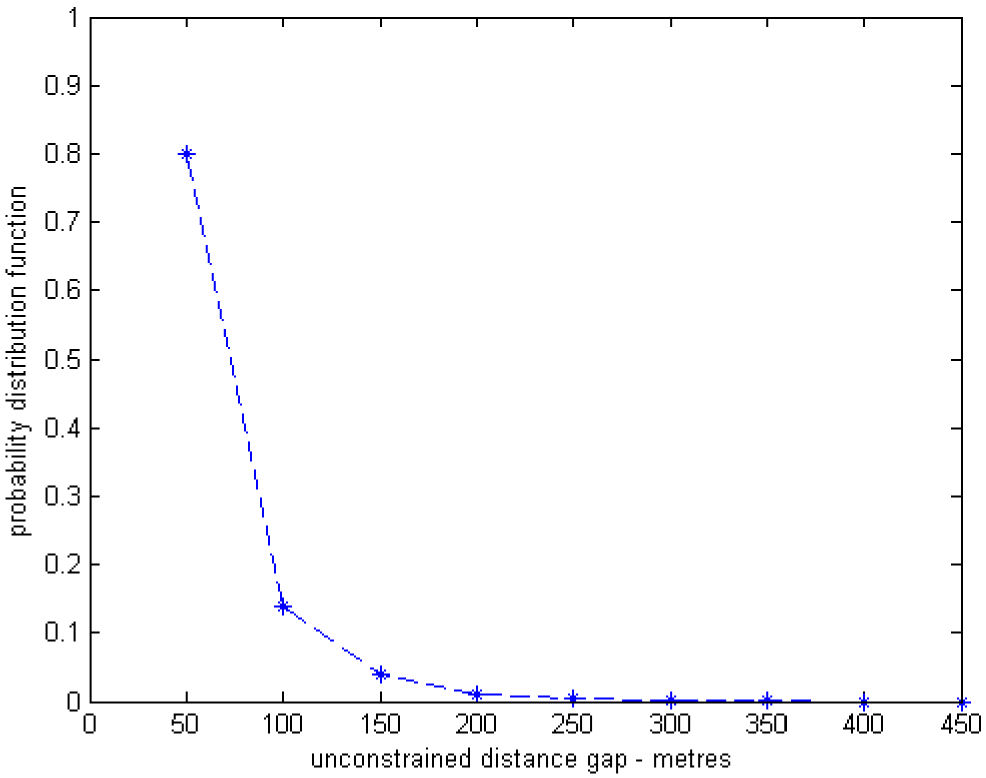


Figure 6.8 Scenario 1 – probability density function for unconstrained distance gap

6.3.2.2. Scenario 2. Constrained distance gap - C_{dg}

If in the first scenario there was no concern about the real utilisation of public transportation networks, in this case several constraints are taken into account.

Arguably, these constraints should make the distance gap to increase even more. Thus, the same failure scenario is considered, in which a vehicle fails in a point of its route, but additionally the following constraints are considered:

- failure of a tram or train carriage, makes the tracks unusable for all the route numbers that share the tracks in the failure area;
- passengers of a failed vehicle cannot be picked up by a tram or train if the failure point is between stations. They can be picked up between stations only by buses. As a consequence if there is no bus route in the area, passengers must walk to the closest rail (tram or train) station;
- trams and trains can only pick up passengers in their designated stations, but they can fail in any waypoint of their routes.

Using the above setup, the probability density function is calculated and displayed in a similar manner as in the first scenario.

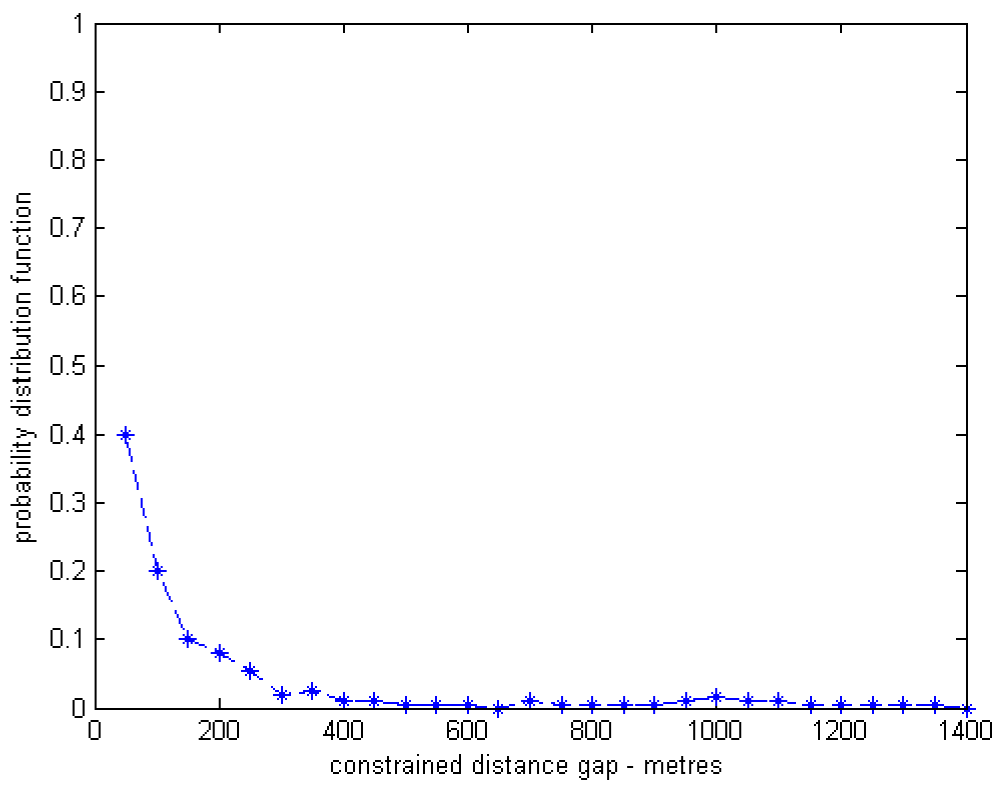


Figure 6.9 Scenario 2 –probability density function for constrained distance gap

Figure 6.9 demonstrates that realistic constraints lead to a considerable increase in distance gap, compared to the unconstrained scenario. Results show that in most cases the passengers will find an alternate transportation mean within 200 metres, which is an acceptable visual range distance. However, in some cases they must walk to as far as 1400 metres away for catching another vehicle. There is clearly a significant difference when compared to the first scenario, in which in the worst cases the walking distance was within 500 metres. A distance of 1400 metres is certainly a significant distance gap, and the existence of such cases raises serious questions about the resilience of transportation system at service level.

In order to further investigate this aspect the time gap (delay) is also considered, for creating complete picture of quality of services associated with the transportation system. This is presented below in the next scenario.

6.3.2.3. Scenario 3. Constrained time gap (delay) - T_g

This scenario is similar to the second scenario, but in addition to distance gap time gap is also considered. Thus, the same type of failures are assumed as in scenarios 1 and 2, with the same constraints as in scenario 2, but the main question is how failures affect the passengers from a delay perspective. Using the timetables for all tram, bus and train routes, the intention is to find how much the passengers have to wait until they can get into another vehicle in the event of a failure. This scenario considers that the affected passengers will walk for a period of time to the closest alternate transportation mean, and then they have to wait in that point for the arrival of the next vehicle. For a more realistic investigation failures are considered to take place in two situations: during the peak and the off-peak hours.

First, walking time from the failure point to the closest available transportation mean is estimated. For this scenario the average speed for human walking was considered to be 5 km/h based on studies of Knoblauch and colleagues (Knoblauch, Pietrucha, & Nitzburg, 1996). The authors found that pedestrian walking speeds range from 5.32 km/h to 5.43 km/h for younger individuals and from 4.51 km/h to 4.75 km/h for older ones.

Second, waiting time to the arrival of the next vehicle is added to walking time (e.g. walking time to the closest eligible tram station plus the time for waiting there for the next carriage belonging to an eligible route number, where eligible situations are as in the second scenario). Then the probability density functions are calculated for the time gap corresponding to failures during peak and off-peak hours.

Figure 6.10 and Figure 6.11 show the results of the investigation. It can be seen that the maximum time gap during peak hours is not greater than 45 minutes, while during off-peak hours it goes to over 60 minutes.

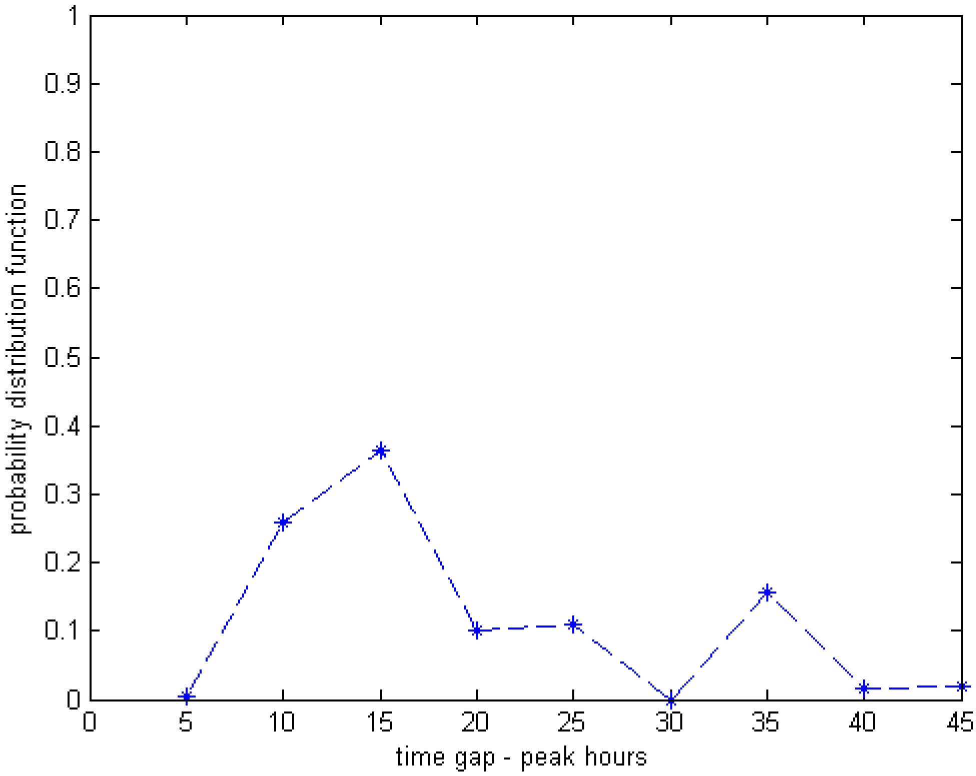


Figure 6.10 Scenario 3 – probability density function for time gap during peak hours

Because during the peak time vehicles are scheduled with a higher frequency than in the off-peak time, this is an expected result providing consistency with real life. However, the question that naturally arises is how acceptable for a passenger is the delay time for the two situations. Results displayed Figure 6.10 indicate that during peak hours the time gap introduced by most of the failures is situated within 10 to 15 minutes. For off-peak hours (Figure 6.11) the highest probability is for delays of 15 to 25 minutes.

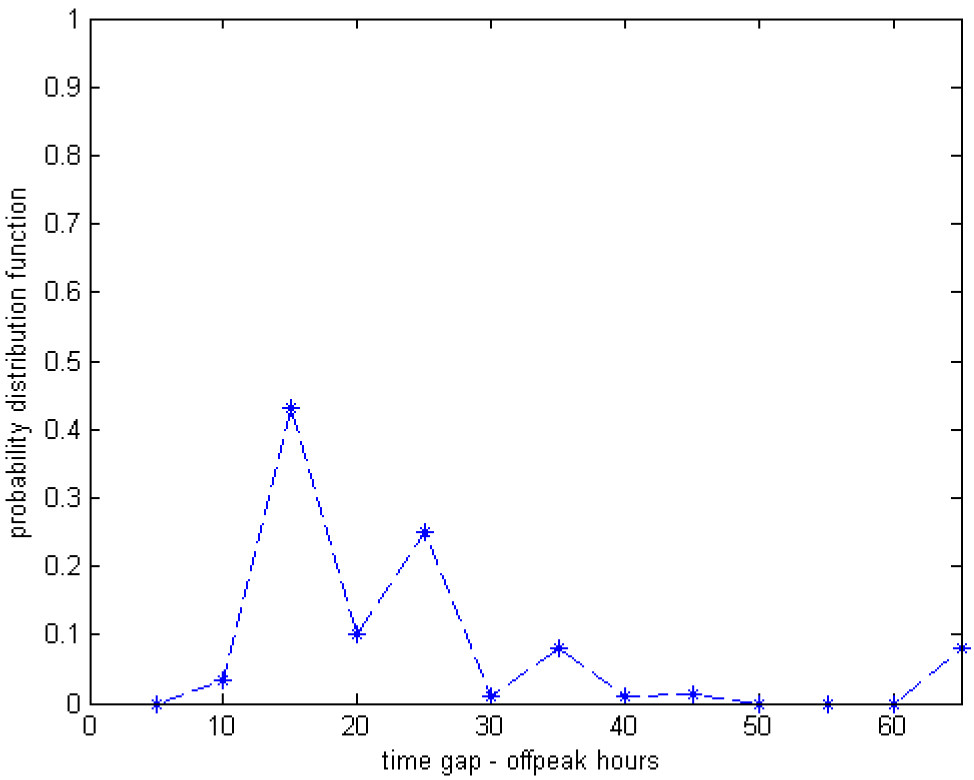


Figure 6.11 Scenario 3 – probability density function for time gap during off-peak hours

The delays may look acceptable and also may appear like a normal result when thinking about the difference in vehicles' frequency for the two time periods (peak and off-peak). But how much would they really affect the passengers in their attempt to reach their destinations in a specific range of time? It can be asked, for example, how important is a delay of 15 minutes in the peak hours for employees who must reach their jobs. Or it can be asked how important is a 25 minutes delay in the off-peak hours, when the passengers are most likely travelling without being time pressured. These are obviously questions that touch the nature of human participants in transportation system. They addresses nevertheless their perception about what is acceptable and what is unacceptable at a certain moment in time. However, this is outside the scope of this section, and also of this chapter. The focus of the investigation is on the gaps generated by failures at service level, for providing an overall view upon the service level resilience.

It should be noted though that in both Figure 6.10 and Figure 6.11, the probability distribution function for a time gap of 35 minutes is higher than for time gaps of 30 and 40 minutes. The reason for this stays nevertheless in the GIS data for Melbourne city. Physical transportation infrastructure, route map and timetables of public transportation generate the results presented in the figure. Since no comparison with other cities has been done, it is impossible to say whether there is a certain reason for higher probability of a 35 minutes time gap or not.

6.4. Resilience assessment at traffic behaviour level

In Section 6.3, the assessment was performed on all transport means and included express roads with speed limits of 100, 80, 70 and 60 km/h and rail networks. For both roads and railways, the physical network layer and the associate public transportation services were considered. Hence, the resultant resilience levels were obtained with a focus on the infrastructure and its passive users (i.e. travellers using public transportation), who do not have a direct influence on the system.

In this section the assessment is performed with a focus on the driver of personal vehicle and on its influence on the system. The investigation takes into account traffic flow of individual vehicles and its variation as result of behavioural pattern of individual car drivers. Thus, only the express roads (speeds of 100, 80, 70 and 60 km/h) and the associated traffic flow are considered, while rail networks and public transport (both road and rail) are neglected.

Two aspects are of main concern for this investigation. First, robustness to waypoint failures is studied with inclusion of traffic. This is different from Section 6.3 where robustness to failures only included topology aspects and public transport aspects. Second, traffic measures are computed assuming a variety of behavioural patterns for the population of drivers that makes use of the existing express road infrastructure.

6.4.1. Investigation of traffic resilience to waypoint failures

Traffic Flow Integrity (F_i). Traffic flow integrity is computed in a similar way as the Topological Integrity, except the measured quantity is the amount of traffic instead of connectivity. For computing the F_i, failures are simulated for each waypoint of the analysed network. The effect of waypoint removal on traffic integrity is observed by computing the number of vehicles remaining unaffected by removal. This number is further divided by the total number of drivers in the network, in order to obtain a sub-unitary ratio. It is likely that some of the waypoints fail and leave the traffic unaffected, while others cause the interruption of certain amount/ratio of traffic. Thus, the traffic flow integrity measures the average traffic survival ratio per network accounting for all single waypoint failures.

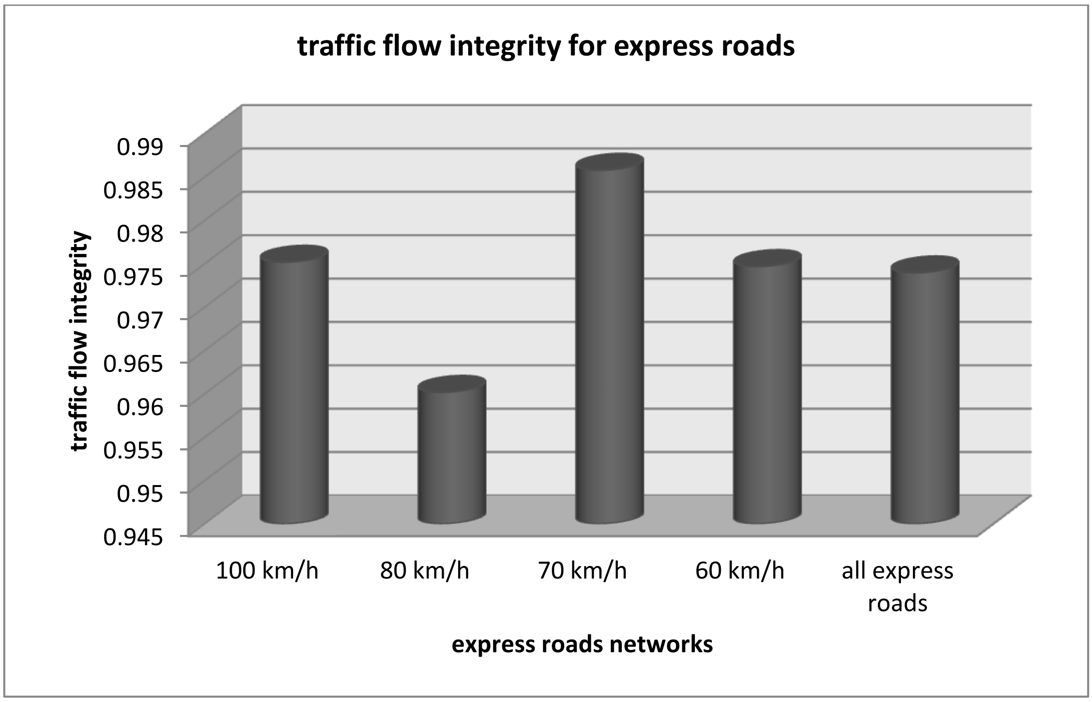


Figure 6.12 Traffic flow integrity for road networks with speed limits of 100, 80, 70 and 60 km/h

Figure 6.12 shows the traffic flow integrity for each express road network, as well as the overall integrity for all networks. All networks display very high traffic integrities, with F_i ratios situated above 0.96. Among them streets with speed limits of 70 km/h have the highest resilience, while those of 80 km/h have the lowest one.

This is further investigated by calculating the probability that removal of a waypoint leaves unaffected (alive) a certain percentage of traffic. Figure 6.13 shows this probability for each of the express road networks, considering a set of five percent-based traffic integrity intervals: <50%, 50-80%, 80-90%, 90-95% and >95%.

Results show again high levels of robustness for all express road networks, with a slight advantage for road network with speed limit of 70 km/h. Though, the differences are not obvious enough to conclude that one of the networks is certainly better than the others. However, it can be concluded that all express road networks in Melbourne appear to be very well designed and to have high resilience to single waypoint failures from a traffic integrity point of view.

From one point of view, the traffic flow integrity can bring in some extent traffic-related information to system planners, as opposed to topological integrity which only shows connectivity-related aspects. In a different view even if traffic integrity is very high, the resultant system performance can be still very low if other traffic measures are considered, such as average speed. This aspect is further investigated in the following paragraphs.

6.4.2. Investigation traffic resilience to collective driver behavioural pattern

Investigation performed in this sub-section uses populations of SoM driver agents with various heterogeneity degrees, as explained in Section 6.2 and using a similar multi-agent setup as in Chapter 5.

The investigation is conducted on a set of populations with different levels of heterogeneity, i.e. normal distribution of personality traits with μ=0 and $\sigma \in (0; 1]$. This set includes a realistic population of drivers with the same distribution of personality traits as the real Australian population: normal distribution with μ=0 and σ=0.7 (Allik & McCrae, 2004; R. R. McCrae & A. Terracciano, 2005; David P. Schmitt, et al., 2007).

For all cases, the traffic performance is assessed with regard to the average speed in different locations of the physical express road networks.

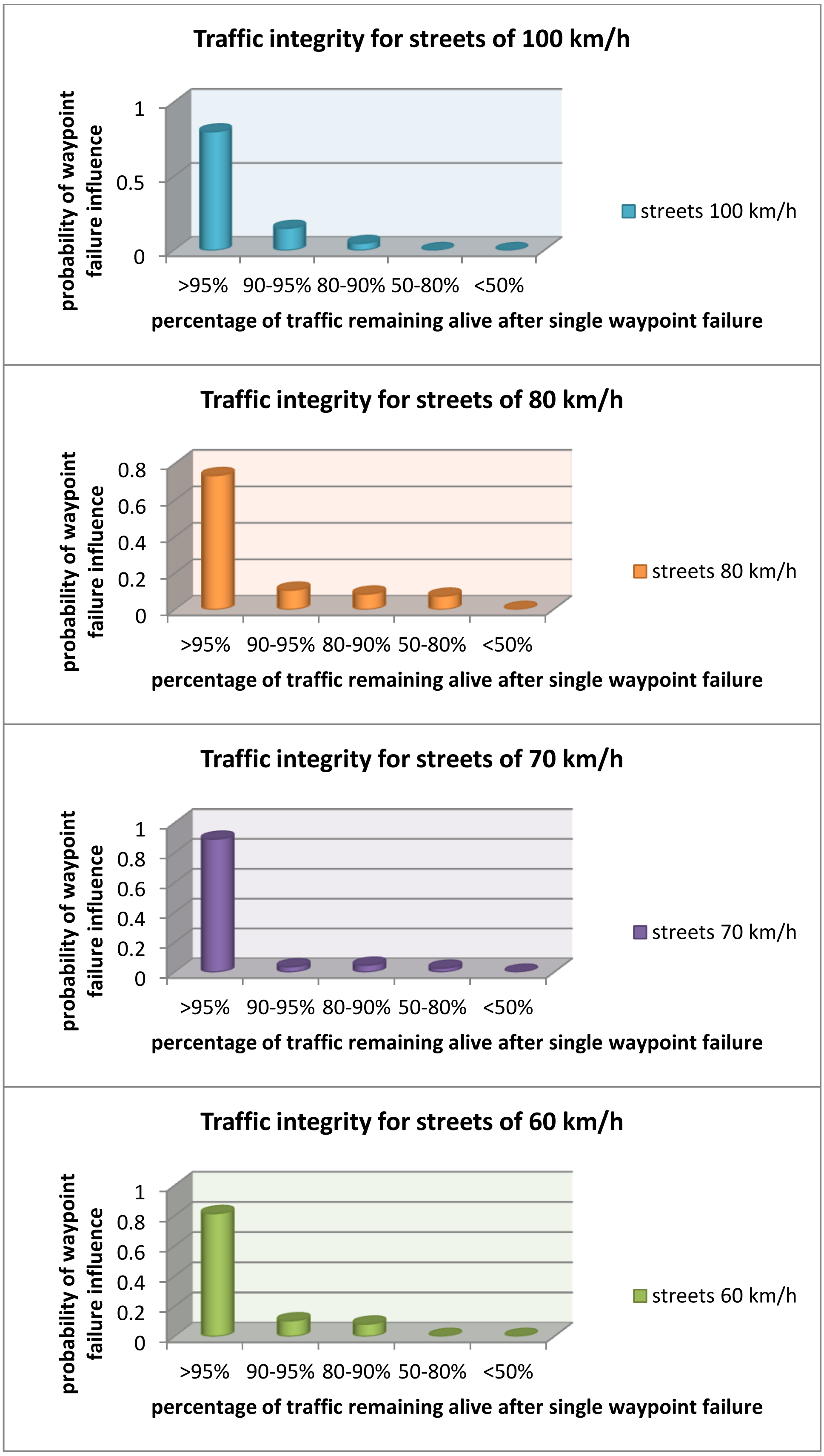


Figure 6.13 Influence of waypoint failure on traffic (probability) for express roads networks

Spatial distribution of average speed (S_v). This measure shows the average speed of vehicles passing through various road segments situated in different locations across the city. Speed sensors placed in each waypoint of GPS maps record instantaneous speeds of the vehicles passing by. The spatially located average speed is then calculated for each sensor over the duration of the whole traffic simulation. However, this measure only has a visual impact, placing coloured dots on the city map. The average speeds are displayed with different colours for each the express road network. For each network, lighter nuances show lower speeds and darker nuances show higher speeds.

Figure 6.14 shows the distribution over the city of average speeds for the realistic population (σ=0.7). Figure 6.15 and Figure 6.16 display the same measure for two extremes: highly homogeneous population (σ=0.1) and highly heterogeneous population (σ=1), respectively.

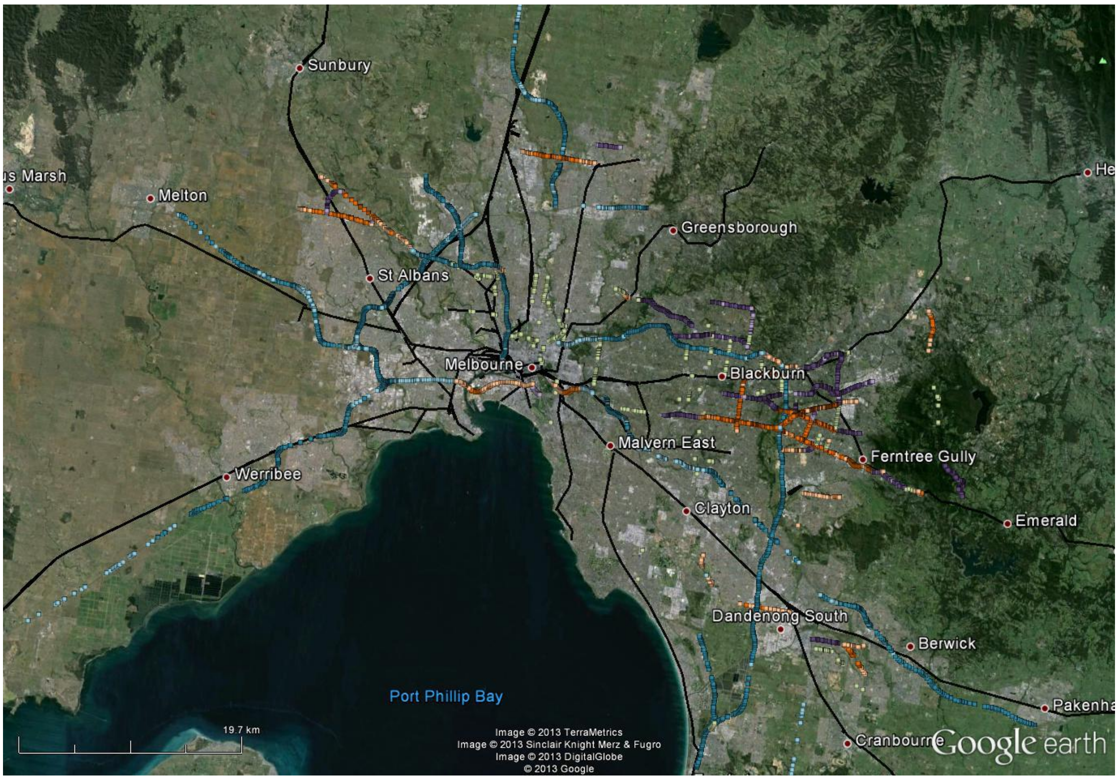


Figure 6.14 Spatial distribution of average speed, σ=0.7 (realistic Australian population): ■ – 60km/h, ■ – 70km/h, ■ – 80km/h, ■ – 100km/h (■ – rail network); darker colours – higher average speeds

Visually, it can be concluded that over all the images higher average speeds for homogeneous population and lower average speeds for heterogeneous populations. The realistic population is situated in between closer to the heterogeneous end of the spectrum. However, such a conclusion could be subjective due to the visual nature of the observation. Thus, this aspect is further investigated by displaying the probability that average speeds fall in certain speed intervals, conveniently chosen for each express network. Figure 6.17, Figure 6.18 and Figure 6.19 show these probabilities and confirm the visual information that average speed is higher for high homogeneity and lower for high heterogeneity. Also, the realistic population is between the two extremes, with the average speed probabilities being very close to those corresponding to highly heterogeneous population.

Figure 6.15 Spatial distribution of average speed, σ=0.1 (low heterogeneity): ■ – 60km/h, ■ – 70km/h, ■ – 80km/h, ■ – 100km/h, ■ – rail network. Darker colours – higher average speeds

In order to get a clearer view on this finding, average speed probability is computed for all range of heterogeneity: $\sigma \in (0; 1]$. Figure 6.20 shows how probability for high speed decreases strongly with higher heterogeneity, while medium speed increases strongly and low speed is fairly steady or decreases slightly. This demonstrates that traffic performance decreases with population heterogeneity, the finding being consistent with results obtained for the artificial road network investigated in Chapter 5.

Figure 6.16 Spatial distribution of average speed, σ=1 (high heterogeneity): ■ – 60km/h, ■ – 70km/h, ■ – 80km/h, ■ – 100km/h (■ – rail network); darker colours – higher average speeds

Average speed per network (N_v). Average speed per network shows for each express road network (100, 80, 70 and 60 km/h) the overall average speed recorded by all speed sensors for all vehicles throughout the simulation.

Figure 6.21 confirms that the average speed over each network decreases when population heterogeneity increases. On one side, this fact is consistent with the results discussed in above paragraphs and also with the results discussed in Chapter 5. However, from a different point of view this demonstrates a dramatic under-utilisation of system resources when driver behaviour is taken into account.

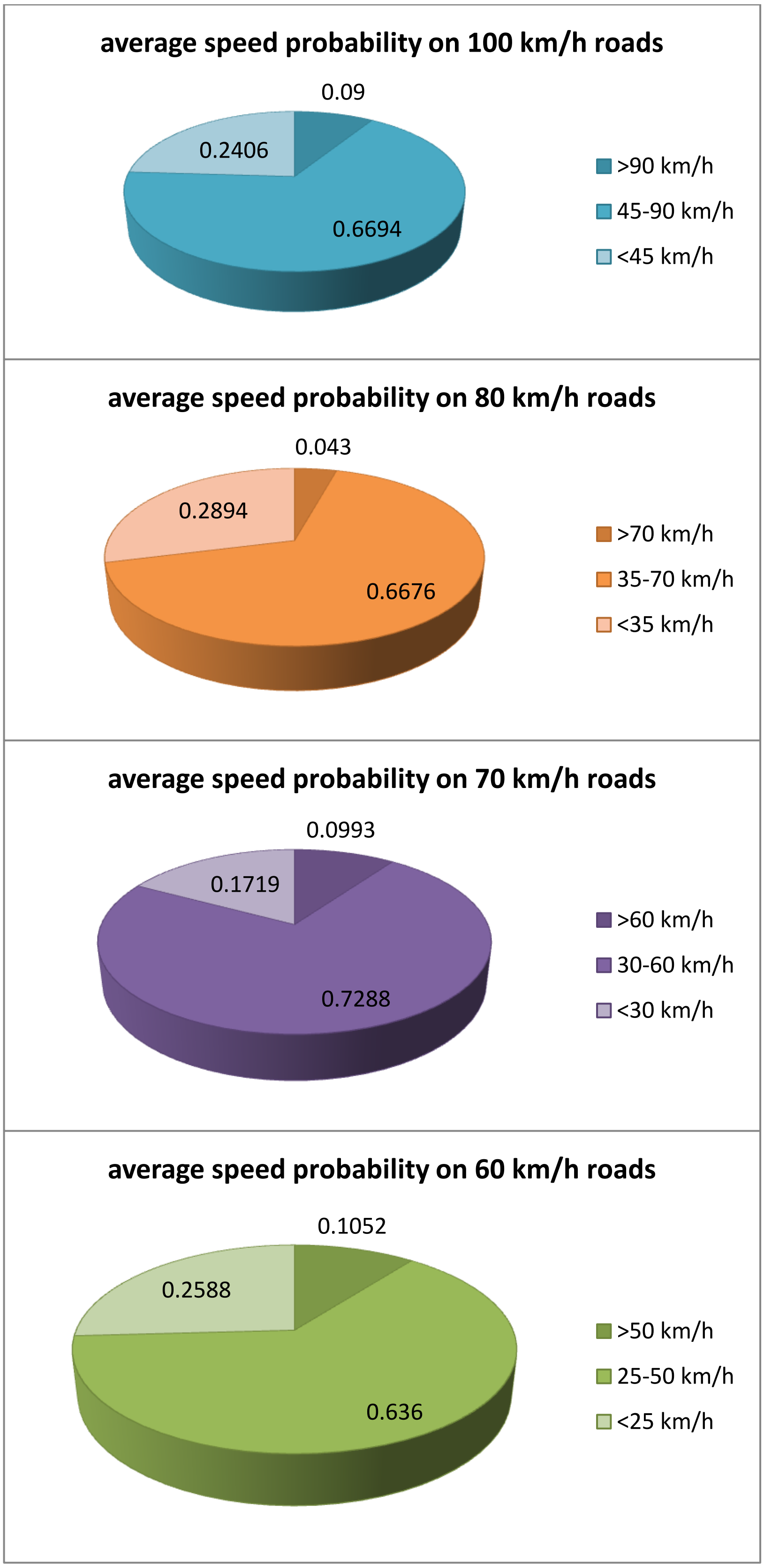


Figure 6.17 Average speed probability on express roads for σ=0.7 (realistic Australian population of drivers)

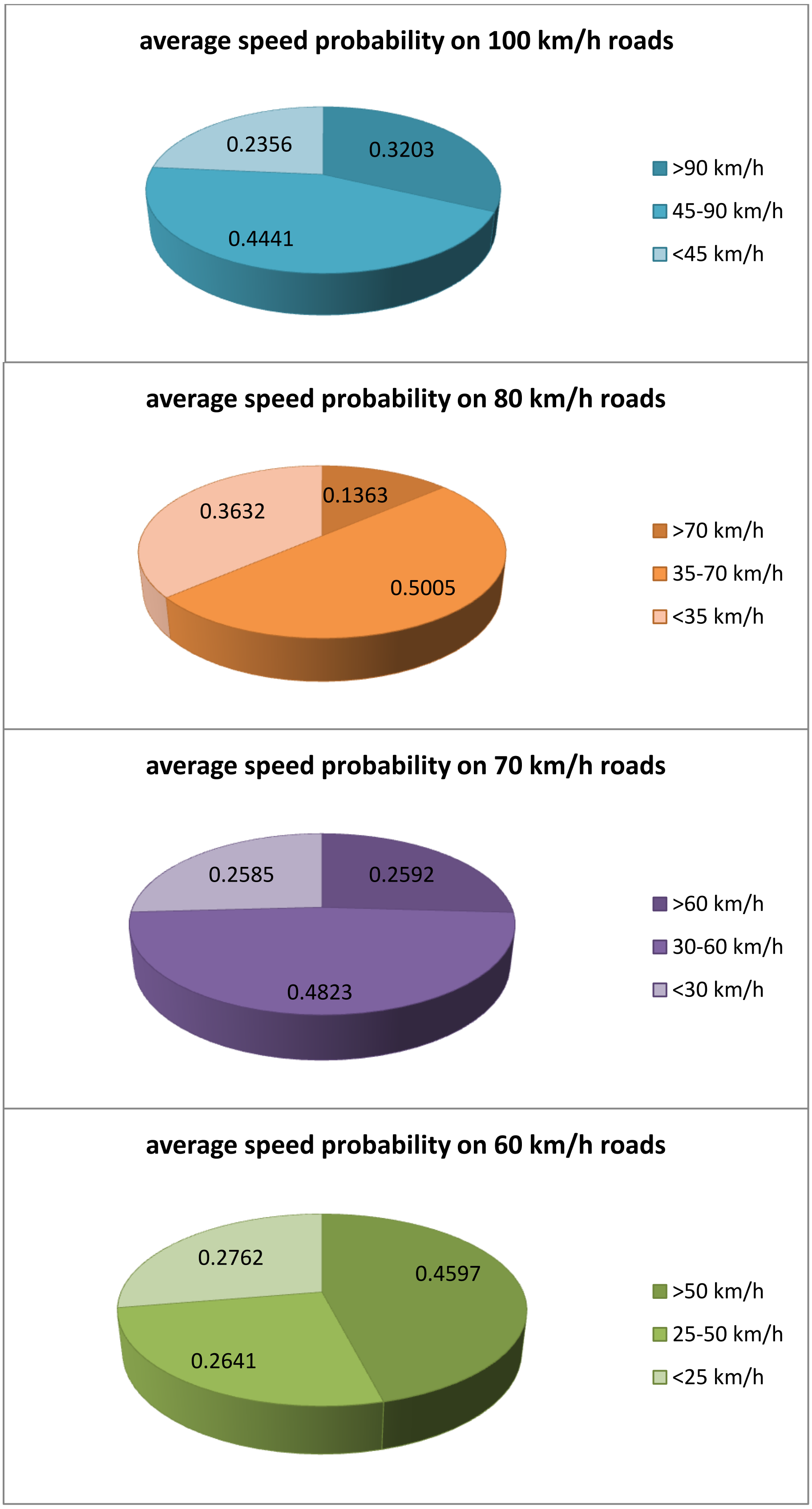


Figure 6.18 Average speed probability on express roads for σ=0.1 (population with very low heterogeneity)

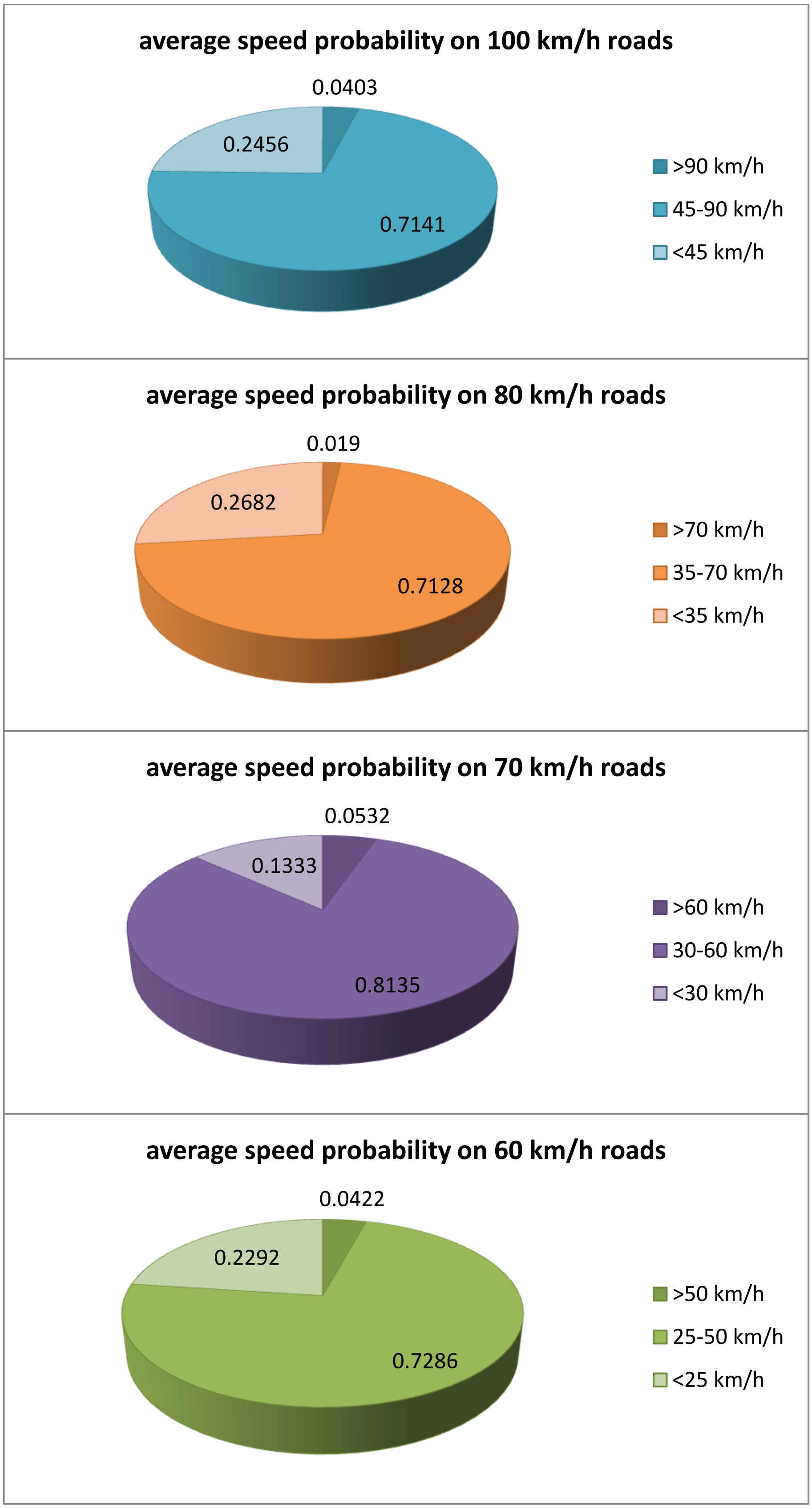


Figure 6.19 Average speed probability on express roads for σ=1 (population with very high heterogeneity)

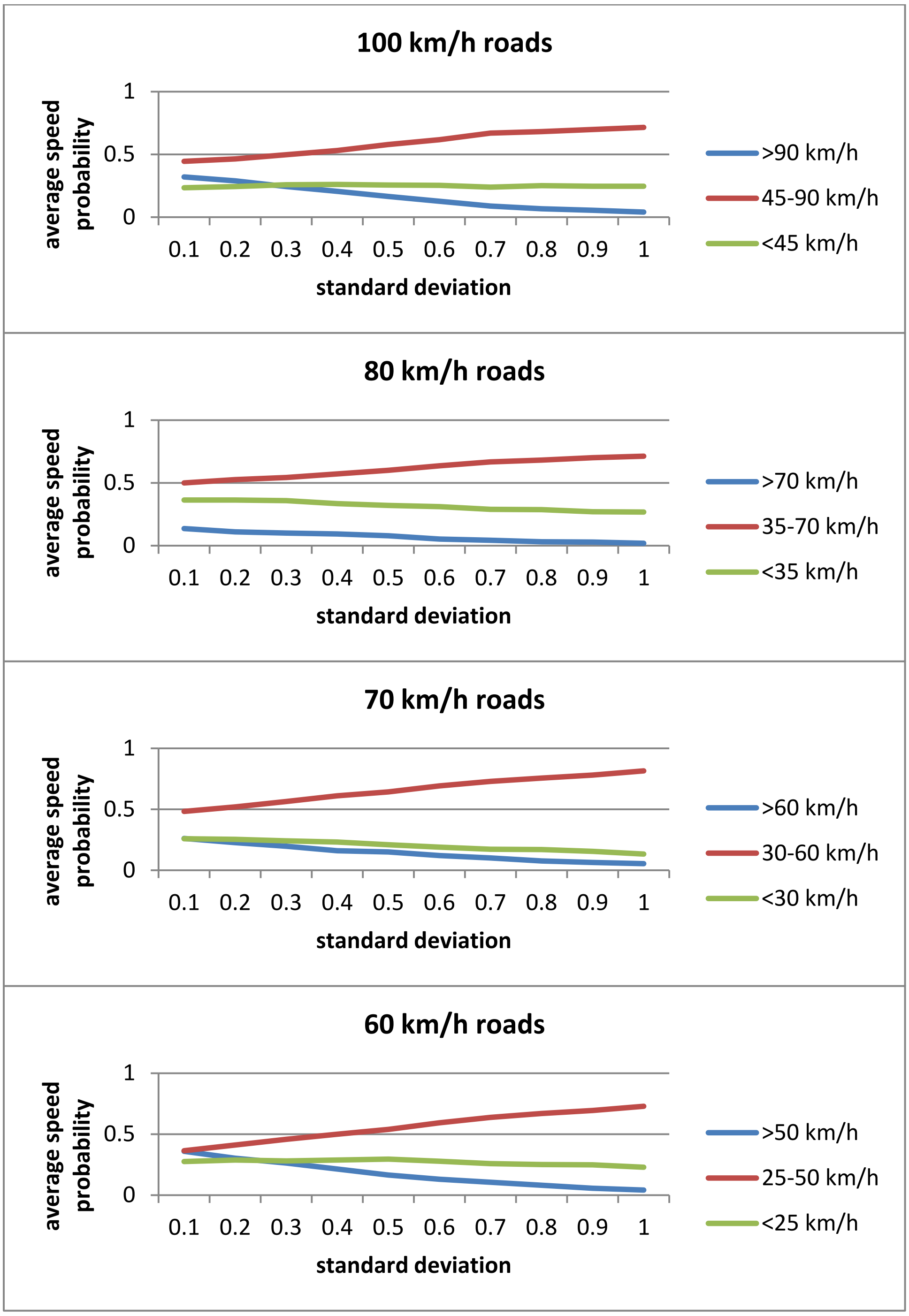


Figure 6.20 Average speed probability per network for the whole range of heterogeneity: $\sigma \in (0;1]$

As an example, it can be clearly seen that on roads with speed limit of 100 km/h the actual average speed for realistic population is only 66 km/h, while for the best case of a highly homogeneous population the average speed is 82 km/h. Thus, despite the infrastructure exists and high speeds are legally and physically possible, system utilisation remains very low, depending dramatically on the behavioural pattern of the local population of drivers.

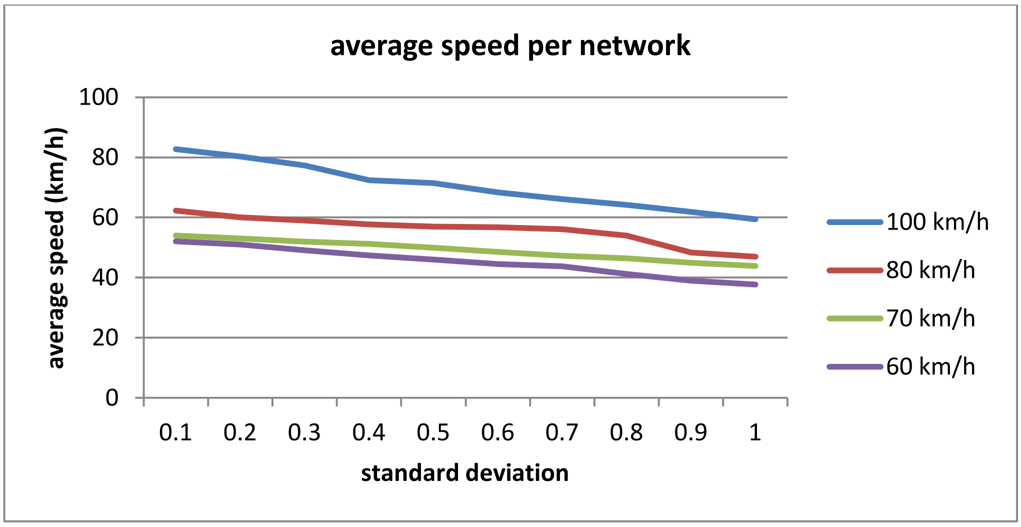


Figure 6.21 Average speed per network for the whole range of heterogeneity: $\sigma \in (0; 1]$

Such a conclusion creates the motivation and the foundation for defining a resilience index for city transportation systems. This index should not only measure infrastructure elements, either at physical or service level, but should also include elements from the behavioural domain. This is described in detail in Section 5 of this chapter.

6.4.3. Comparison with real traffic data

The simulated results of the behavioural assessment are compared with real road network data are extracted from the maps of Melbourne area (courtesy of Google Earth and Open Street Maps) and used with suburb population distribution data taken from Victorian Government population bulletin (Government, 2010) and road traffic statistics from Australian road traffic association (Austroads, 2013) and Victorian (Melbourne) road traffic authority (VicRoads, 2012).

The aggregate average speed over the inner and outer freeway roads of Melbourne (Figure 6.22) is calculated and compared to the real average speed reported by road traffic authorities. The real population data, the real road network and the recorded average speed are those corresponding to year 2010 (Figure 6.23).

Figure 6.22 Melbourne inner and outer freeways network: ■ –roads with speed limits of 100km/h, ■ – rail network.

Figure 6.24 shows the simulated average speeds for various degrees of heterogeneity, including a standard deviation of 0.7, which corresponds to the assumed realistic population of Melbourne. From one point of view, results show that traffic performance decreases with the population heterogeneity, confirming in a realistic context the results obtained for a single road segment. From a different point of view, the average speed generated by the realistic population of Melbourne (71.41 km/h) is consistent with the real average speed (76 km/h) recorded by traffic authorities, with a 6.03% deviation from real data. This demonstrates that the proposed SoM agent can be successfully used in a realistic context, generating traffic outcomes very close to the recorded traffic data. However, a deviation exists, which can be the result of several factors. First, the road network was obtained by accessing Open Street Map website (www.openstreetmap.org) and exporting the area of interest. Since Open Street

Map is an open source system updated by volunteer contribution, certain differences from real road maps may exist, potentially introducing a subsequent amount of inaccuracy. Second, the population of Melbourne was considered as having similar statistical distribution of personality features to that of the whole Australian population, fact that could introduce as well a certain degree of inaccuracy.

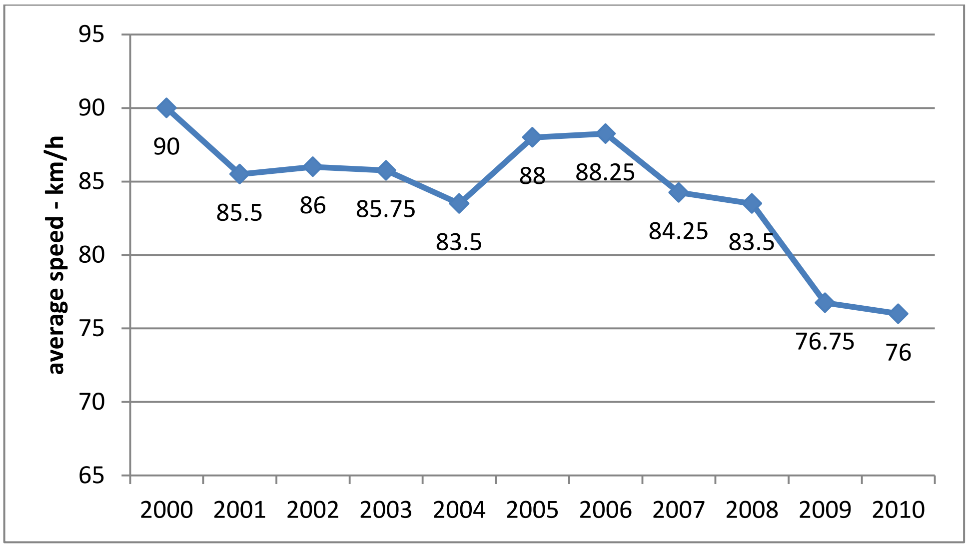


Figure 6.23 Yearly aggregate average speed for Melbourne freeway roads. Data from (Austroads, 2013) and (VicRoads, 2012).

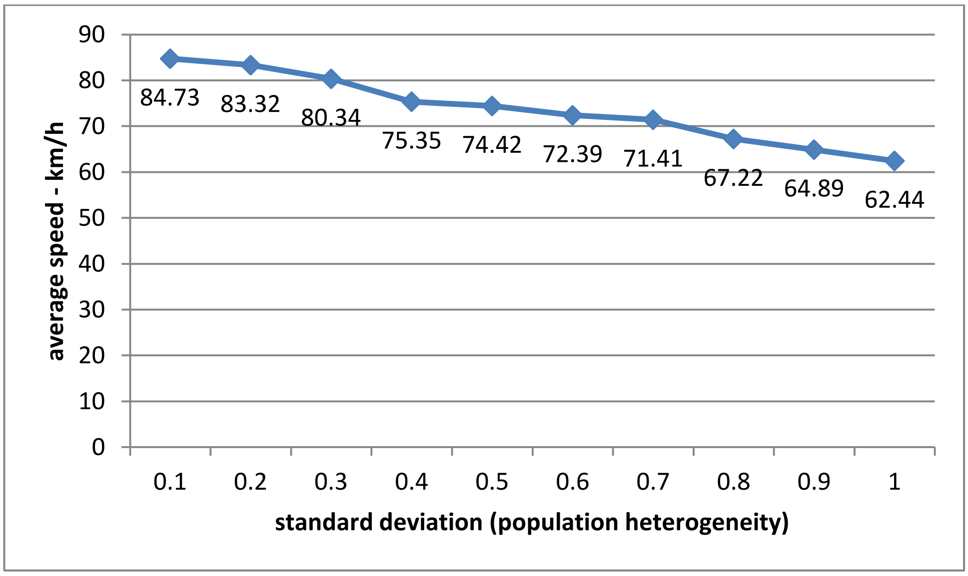


Figure 6.24 Aggregate average speed for Melbourne freeways roads. SoM driver populations with various degrees of heterogeneity.

6.5. City transportation resilience index

In this section an objective measure for system level city transportation resilience is proposed as a way to include all aspects discussed in this chapter, both infrastructure and behaviour related.

6.5.1. Definition

Definition of such an index should start from the idea that a certain deviation of actual performance from the expected performance exists and is measurable. Hence, the resultant index is a ratio showing either the actual level of performance or the deviation from ideal in relation with the ideal (planned) performance level. Also, the definition must take into account multiple aspects of the transportation system such as physical infrastructure, services and behaviour. Thus, the overall resilience index must be presented as a mean value of all partial indexes. Equation (6.1) shows the general form of the proposed resilience index.

In particular for this thesis, three partial indexes are considered, equation (6.2), corresponding to the fundamental domains investigated in this chapter. The infrastructure domain accounts for a physical layer index (I_p) and a service layer index (I_s), while the behavioural domain accounts for a behavioural index (I_b).

$$\boldsymbol{I_r = \frac{1}{n}\sum_{k=1}^{n} I_k} \tag{6.1}$$

$$\boldsymbol{I_r = \frac{1}{3}(I_p + I_s + I_b)} \tag{6.2}$$

6.5.2. Partial index for infrastructure resilience – physical layer (I_p)

Physical layer partial index is calculated based on the ratio of waypoints with topological integrity greater than 2. Topological integrity, presented in Subsection 6.3.1 as a measure of robustness/resilience to waypoints failures, represents the most comprehensive measure. It implicitly contains information about all connectivity measures, and also represents the foundation for computing all measures for spatial distribution of risk.

I_p is computed so that it shows the deviation from ideal, where the ideal situation (high robustness) means no waypoint with T_i>2. Figure 6.25 displays the ratio of waypoints with T_i>2 for all networks considered for assessment in this chapter. The overall ratio (Ti_r=0.1355) is calculated as the average of all networks ratios, as in equation (6.3):

$$\boldsymbol{T_{i_r}} = \frac{\mathbf{1}}{\mathbf{6}}\sum_{\boldsymbol{i=1}}^{\mathbf{6}}\frac{\boldsymbol{n_i}}{\boldsymbol{N_i}} \tag{6.3}$$

where n_i and N_i are the number of waypoints with $Ti>2$ and the total number of waypoints, respectively, for all six networks considered for assessment.

Thus, the physical level partial resilience index is calculated as in equation (6.4), and the resultant value is $I_p=0.8645$.

$$\boldsymbol{I_p} = \mathbf{1} - \boldsymbol{T_{i_r}} \tag{6.4}$$

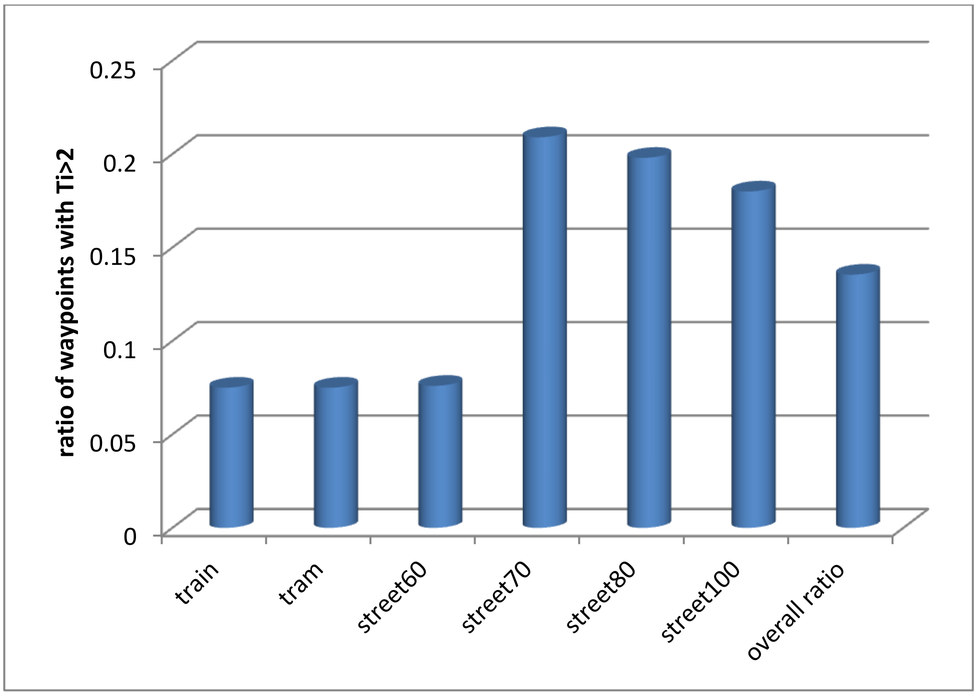


Figure 6.25 Physical level – ratios of waypoints with T_i>2

6.5.3. Partial index for infrastructure resilience – service layer (I_s)

Service layer partial index is calculated based on the ratio of waypoints with time gap greater than 15 minutes for both peak and off-peak time of public transportation means (both road and rail). Time gap was discussed in Subsection 6.3.2 as a measure of robustness/resilience to waypoints failures in relation with quality of services offered by public transport networks. It is a representative measure for service level resilience since it includes implicitly all scenarios discussed in that sub-section.

I_s is computed so that it shows the deviation from ideal, where the ideal situation (high robustness) means no waypoint with T_g>15 minutes. Figure 6.26Figure 6.25 shows the ratio of waypoints with T_g>15 minutes for all public transportation networks and means, for peak-time and off-peak time. The overall ratio (Tg_r=0.4718) is calculated as in equation (6.5), where n_i and N_i are the number of waypoints with T_g>15 and the total number of waypoints, respectively, for peak and off-peak time. Thus, the service level partial resilience index is calculated as in equation (6.6), and the resultant value is I_s=0.5282.

$$T_{g_r} = \frac{1}{2}\sum_{i=1}^{2}\frac{n_i}{N_i} \tag{6.5}$$

$$I_s = 1 - T_{g_r} \tag{6.6}$$

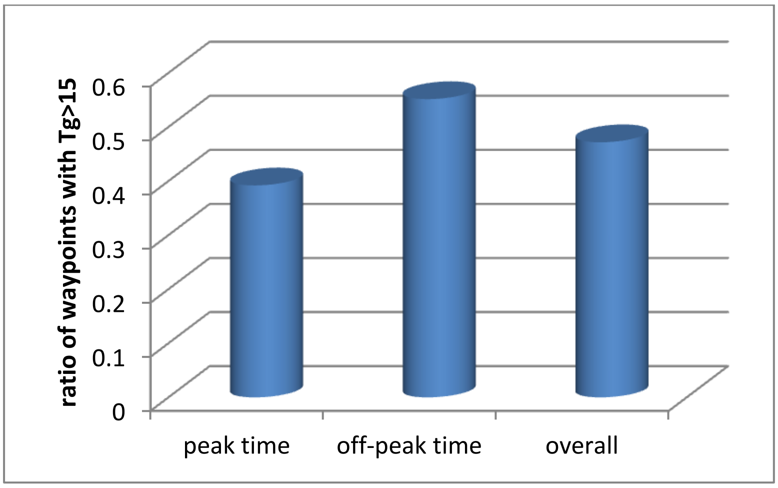


Figure 6.26 Service level – ratios of waypoints with T_g>15, for peak and off-peak time

6.5.4. Partial index for behavioural resilience

Behavioural level partial index is calculated based on the average speed per network (N_v) as a ratio showing the deviation of actual speed from maximum speed limit of the respective network.

I_b is computed to show the deviation from ideal, where the ideal situation (high resilience) means average speeds equal to maximum legal speed. However, this partial index has different values for different types of driver populations. Figure 6.27Figure 6.26Figure 6.25 shows for populations of all range of heterogeneity the average speed ratio for all express roads networks, including the behavioural resilience index (I_b) calculated as in equation (6.7):

$$I_b = \frac{1}{4}\sum_{i=1}^{4} \frac{N_{v_i}}{v_{max_i}} \tag{6.7}$$

where N_{vi} and v_{maxi} are the average speed and the maximum legal speed, respectively, for each type of express roads network.

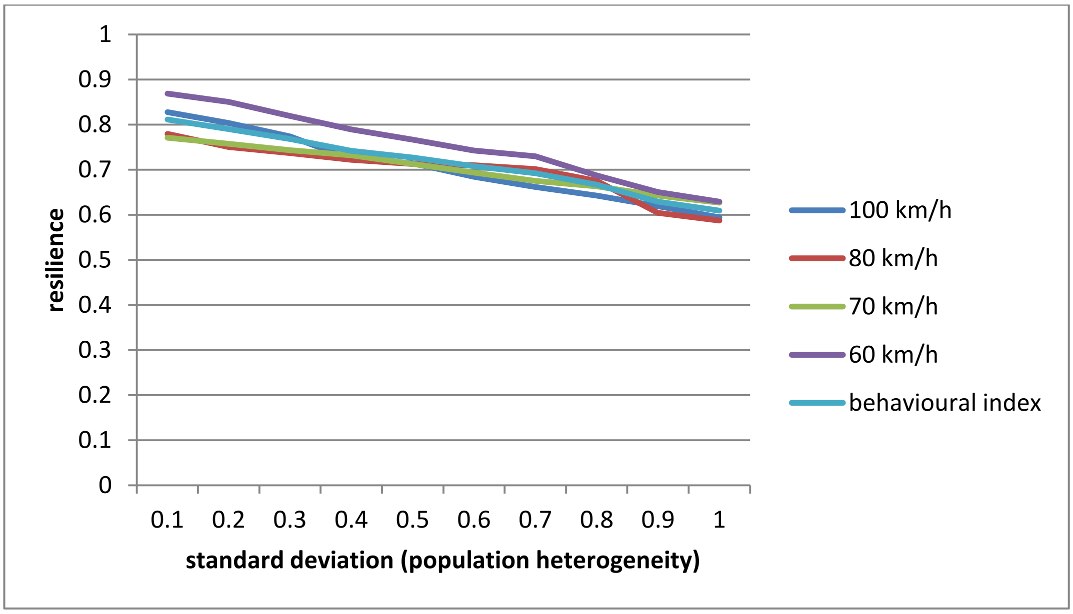


Figure 6.27 Behavioural level – average speed ratios for express roads networks and behavioural level partial resilience index (I_b) for populations with various homogeneity levels

6.5.5. The city resilience index

City resilience index is calculated according to equation (6.2) as the average (mean) of partial indexes. However, while infrastructure related partial indexes are unique for the given infrastructure, the behavioural index varies depending on population behaviour. This variation is shown in Figure 6.28, which depicts possible values of the city resilience index assuming different behavioural patterns of the local populations of drivers. It is clearly shown that overall the resilience of transportation system decreases with the heterogeneity.

Also, the resilience index of Melbourne for an assumed realistic population (R. R. McCrae & A. Terracciano, 2005) is I_r*=0.695*, indicating an significant deviation of transportation performance from the ideal performance level.

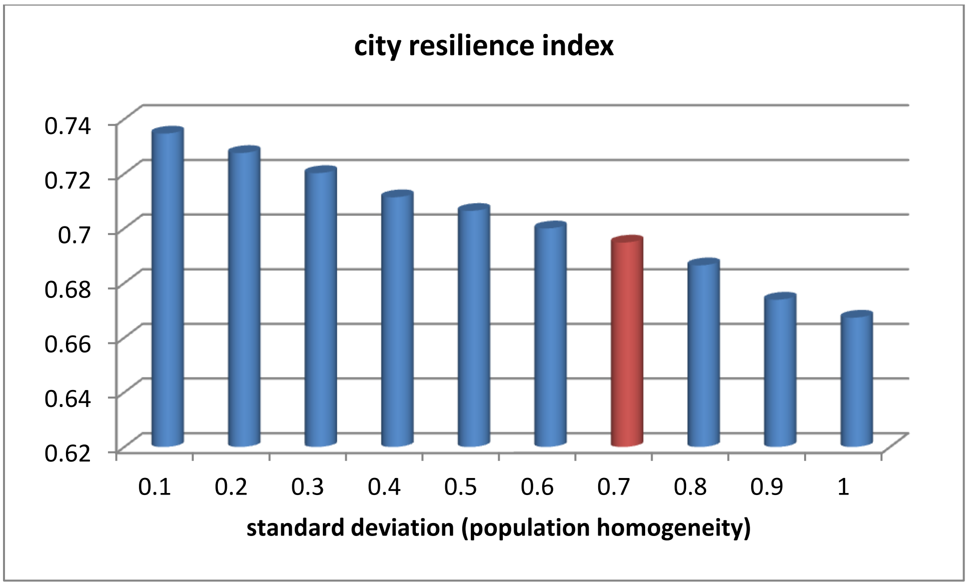


Figure 6.28 City resilience index for various populations with various behavioural pattern (heterogeneity): ■ – Melbourne realistic population, ■ – Melbourne other populations

6.6. Discussion

This chapter presented a case study on Melbourne transportation system. The city-scale investigation aimed to show that resilience of ground transport systems depends dramatically on the behaviour of population that makes use of it, at least in the same extent as on the quality of infrastructure. Several aspects have been demonstrated throughout the chapter, as follows.

First, it was demonstrated that the driver agent instantiation of SoM agent architecture, previously tested in an artificial environment, can be successfully used in real world contexts for investigating complex issues. One such issue is the behavioural aspect involved in resilience of road transport systems of real geographical areas.

Second, from a transportation perspective, the chapter provided valuable insights on the resilience of Melbourne transportation system at two fundamental levels: infrastructure and behaviour. Infrastructure is what is offered to local population and investigation in this direction considers system users as passive entities that do not influence the system. Hence, the assessment is performed on physical and service layers, where physical assessment treated connectivity issues and service assessment treated public transport issues. Behavioural level treats system users as active entities which influence the system through their behaviour. Hence, the focus was on the traffic performance generated by individual car drivers through the resultant population-level behavioural pattern. Results demonstrated that investigation of behavioural aspects is a *sine qua non* condition for performing pertinent and complete resilience assessments of transportation systems.

Third, the chapter in itself can be seen as presenting a framework for assessing the resilience of city-scale ground transportation systems. The proposed framework concludes with an index-based measurement of city transportation resilience, which measures the deviation from ideal (planned) level of performance. Furthermore, this framework is a cost-effective, versatile and portable solution. It can be used for any geographical area for which assessment data is available through local or global sources, such as GIS systems, transportation authorities, government reports, population studies and statistics, etc.

With regard to the research questions of this thesis, this chapter treated partially the research question number 3.

Chapter 7. Conclusions

This thesis introduced a novel non-hierarchical hybrid cognitive agent architecture as an alternative to existing hierarchically layered hybrid architectures. The proposed agent architecture is based on the "Society of Mind" theory on human mind and cognition and is intended to demonstrate that a non-hierarchical approach can be successfully used for modelling complex behaviours.

In order to demonstrate its viability the proposed architecture was tested in a cognitively demanding environment, as a driver agent in traffic behaviour context. Throughout the thesis it was shown that the SoM agent architecture, through its driver agent instantiation, had a great power of representation of human cognitive-affective processes. It was able to implement and handle the whole range of human cognitive capabilities – affective, rational, and reactive – while keeping a slender logical and computational construction. Also, it provided native support for inclusion of behavioural biases and propensities, both innate and acquired. These aspects are summarised in the next session.

7.1. Summary of contributions

Overall, the main findings and results of the research work presented in this thesis can be summarised as follows:

- **First**, the Society of Mind has been identified as an appropriate starting point for generating a non-hierarchical cognitive agent architecture. The agent architecture was proposed as a result of a thorough investigation of Minsky's theory on human mind and cognition. Then, based on the findings, the general *Society of Mind* agent architecture was described in detail. The description treated the internal dynamics of SoM agent and also proposed the general methodology – as well as the formalism – for potential implementations. This responded to research question 1, and was treated in Chapter 3.

- **Second**, behaviours on road traffic were used as a relevant context for evaluating the general SoM agent architecture. The architecture was instantiated as a car driver agent. The SoM driver agent was tested in an individual context in various car-following situations in order to demonstrate that such an implementation can produce a wide range of driving behaviours. The internal decision making mechanism and the effect it actuates in traffic conditions were recorded and thoroughly analysed. The agent showed internal consistency, with the internal dynamics adapting to a variety of traffic contexts. The decision-making mechanism actuated the expected actions from both agent perspective and road traffic perspective. Then, the SoM driver agent was further tested in a multi-agent setup, in order to understand the emergence of collective driver behaviour and to study the effect of various behavioural patterns on road traffic performance. Populations of SoM driver agents initialised with various personality patterns were simulated in an artificially generated traffic environment. The simulation outcomes were recorded and discussed from both agent perspective and traffic performance perspective. The populations of driver agents demonstrated collective behaviours that followed the internal dynamics and adaptation patterns of individual agents. It was shown that interaction of individual agents in a multi-agent setup can produce consistent and expected aggregation towards a population level behaviour. This demonstrated that SoM driver agents can be successfully used in complex road traffic simulations. Also, from a road traffic perspective the populations of driver agents produced the expected values for a variety of traffic performance metrics. Thus, it was demonstrated that SoM driver agents can be used successfully for in-depth analyses of road traffic behaviour and psychology. This responded to research question 2, and was treated in Chapter 4 and Chapter 5.
- **Third**, the SoM driver agent was used in a full-sized assessment of real-life ground transportation systems, in a case study on city of Melbourne. It was shown that the Society of Mind agent architecture can become a useful tool for investigating the behavioural aspects involved in performance of road

traffic networks. In this context, the focus shifted towards the assessment of transportation system resilience. The SoM agent was only one investigation tool which added to a more comprehensive set of tools and metrics in an attempt to bring together both infrastructure and behavioural aspects. The investigation showed that the sole use of infrastructure related resilience metrics was insufficient for capturing in a realistic manner system resilience. The addition of behavioural assessment revealed a more useful and complete picture of the resilience landscape. Moreover, this addition was possible through the use of SoM driver agents. This responded to research question 3, and was treated in Chapter 6.

The three main contributions presented above answered to the three research sub-questions of this thesis. Through this, the main research question was entirely and thoroughly addressed. The thesis presented and demonstrated the usability of an agent architecture that allows representation of a wide range of human cognitive capabilities and behavioural patterns without using hierarchical control frameworks.

7.2. Future work

The research work presented in this thesis can be further extended in three main directions, presented below in a random order:

Transportation. Further work in this direction can expand the utilisation of SoM driver agents in a wider range of traffic behaviour and psychology investigations. On one side, the driver agent can be used in more complex city scale traffic scenarios, in which the behavioural component consists of more than one type of driver. Arguably, more realistic assessment outcomes could be obtained if drivers were split into personal car drivers, truck drivers, bus drivers, etc. On a different side, driver agents can be used at smaller scale in purely microscopic and sub-microscopic simulations. These can investigate driver actions and reactions to cognitive stimuli in contexts such as accident investigation and prevention, or automotive research on safety systems.

SoM agent applications. Perhaps the main body of further work, generated by the very existence of this thesis, is on expanding the application range for the proposed architecture. The purpose would be to evaluate the SoM agent architecture in various fields of activity and to position it as a viable architecture for general intelligence. With the appropriate changes in the implementation, several major directions can be investigated:

- business/governance
    - modelling behaviour of economic agents, bidders, buyers, competitors, markets
    - marketing research, product management
- military domain
    - modelling and simulation of military operations, both tactical and strategic
    - modelling and simulation of battlefield psychology
- decision-making
    - modelling of affective decision making and heuristic decision-making
- artificial life
    - modelling and simulation of artificial societies
    - autonomous systems
    - "sugarscape" domains for growing and investigating alternate social evolutionary branches

Development of hybrid cognitive architectures. Further work should be also performed on improving the proposed SoM agent architecture. This can be done either by including more elements from Minsky's theory, which were neglected in the present study, or by using valuable concepts and elements from other existing or emerging cognitive theories.

Appendix 1: RRA SoM agent - JAVA class structure

```
class Environment
{
	double[] _x;
	double[] _y;

	public Environment(double[] x, double y)
	{
		this._x = x;
		this._y = y;
	}
}

class SoMagent
{
	public double[] _personality;
	double[] _x;
	public double[] _y;
	public double[] _xAff;
	public double[] _xRat;
	public double[] _xRct;
	public AffectiveAgency affAgency;
	public ReactiveAgency rctAgency;
	public RationalAgency ratAgency;
	public Bidding bidding;
	public KLine kline;

	public SoMagent(double[] x)
	{
		this._x = x;
		setAgencyInputs();
	}

	void setAgencyInputs()
	{
		this._xAff = affInputSelector();
		this._xRat = ratInputSelector();
		this._xRct = rctInputSelector();
	}
	public void setAgencyOutputs()
	{
		this._y = actuate(  selectedAction( bidding.getBidWinner ) );
	}
}

interface Agency
{
	double getStrength();
	void stateUpdate();
```

```
        void setActions();
        void calculateStrength();
}

class AffectiveAgency implements Agency // sau extends Agency daca e clasa
{
        private double _strength;
        private double[] _emotions;
        private SoMagent _agent;
        private double[] _xAff;
        private AffectiveActionsSet _actions;
        class AffectiveActionsSet
        {
                // define set of affective actions
        }

        public AffectiveAgency(double[] emotions, SoMagent agent)
        {
                this._emotions = emotions;
                this._agent = agent;
                this._strength = 0;
                this._xAff = agent._xAff;
        }

        public void stateUpdate()
        {
                this._emotions = fupdate( _agent._personality, _emotions, _xAff );
        }

        public void setActions()
        {
                this._actions = f( _agent._personality, _emotions, _xAff );
        }

        public AffectiveActionsSet AffectiveKLine()
        {
                KLine.SelectAffActions( _xAff, _actions );
        }

        public void calculateStrength()
        {
                this._strength = fstrength( _agent._personality, _emotions, _xAff );
        }

        public double getStrength();
        {
                return _strength;
        }
}

class RationalAgency implements Agency
{
        private double _strength;
        private Rational _rationals;
        private SoMagent _agent;
        private double[] _xRat;
        private RationalActionsSet _actions;
```

```
        class RationalActionsSet
        {
                // define set of rational actions
        }

        public RationalAgency(Rational rationals, SoMagent agent)
        {
                this._rationals = rationals;
                this._agent = agent;
                this._strength = 0;
                this._xRat = Environment._xRat;
        }

        public void stateUpdate()
        {
                this._rationals = fupdate( _agent._personality, _rationals, _xRat );
        }

        public void setActions()
        {
                this._actions = f( _agent._personality, _rationals, _xRat );
        }

        public RationalActionsSet RationalKLine()
        {
                KLine.SelectRatActions( _xRat, _actions );
        }

        public void calculateStrength()
        {
                this._strength = fstrength( _agent._personality, _rationals, _xRat );
        }

        public double getStrength();
        {
                return _strength;
        }

}

class ReactiveAgency
{
        private double _strength;
        private Reactive _reactives;
        private SoMagent _agent;
        private double[] _xRct;
        private ReactiveActionsSet _actions;
        class ReactiveActionsSet
        {
                // define set of rational actions
        }

        public ReactiveAgency(Reactive reactives, SoMagent agent)
        {
                this._reactives = reactives;
                this._agent = agent;
                this._strength = 0;
```

```
                this._xRct = Environment._xRct;
        }

        public void stateUpdate()
        {
                this._reactives = fupdate( _agent._personality, _reactives, _xRct );
        }

        public void setActions()
        {
                this._actions = f( _agent._personality, _reactives, _xRct );
        }

        public ReactiveActionsSet ReactiveKLine()
        {
                KLine.SelectRctActions( _xRct, _actions );
        }

        public void calculateStrength()
        {
                this._strength = fstrength( _agent._personality, _reactives, _xRct );
        }

        public double getStrength();
        {
                return _strength;
        }
}

class Bidding
{
        double strengthAff;
        double strengthRat;
        double strengthRct;

        public Bidding( double strengthAff, double strengthRat, double strengthRct )
        {
                this.strengthAff = strengthAff;
                this.strengthRat = strengthRat;
                this.strengthRct = strengthRct;
        }

        public double getBidWinner()
        {
                return biddingStrategy( strengthAff, strengthRat, strengthRct );
        }
}

class KLine
{
        public static AffectiveActionsSet SelectAffActions( double[] xAff, AffectiveActionsSet
actions )
        {
                return selectedAffActions;
        }
```

```
	public static RationalActionsSet SelectRatActions( double[] xRat, RationalActionsSet actions )
	{
		return selectedRatActions;
	}

	public static ReactiveActionsSet SelectRctActions( double[] xRct, ReactiveActionsSet actions )
	{
		return selectedRctActions;
	}
}
```

References

Albert, R., Jeong, H., & Barabasi, A. L. (2000). Error and attack tolerance of complex networks. *Nature, 406*, 378-382.

Allik, J., & McCrae, R. R. (2004). Towards a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35*, 13-28.

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355-365.

Arnett, J. J., Offer, D., & Fine, M. A. (1997). Reckless driving in adolescence: [`]State' and [`]trait' factors. *Accident Analysis & Prevention, 29*, 57-63.

Austroads. (2013). Weighted aggregate speed on a representative sample of arterial roads and freeways in major cities. In *National Performance Indicators*. Australia: Austroads.

Bandura, A. (1999). A social cognitive theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality (2nd ed., pp. 154-196)*. New York: Guilford Publications.

Barabási, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications, 272*, 173-187.

Barcelo, J., Ferrer, J. L., García, D., Grau, R., Florian, M., Chabini, I., & Le Saux, E. (1998). Microscopic traffic simulation for ATT systems analysis, a parralel computing version. In.

Barnes, A., & Thagard, P. (1996). Emotional decisions. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society. Erlbaum*, 426-429.

Bell, M. G. H., & Iida, Y. (1997). *Transportation network analysis*. Chichester: John Wiley & Sons.

Berdica, K. (2002). Transport Policy. *An introduction to road vulnerability: what has been done, is done and should be done, 9*, 117.

Blum, G. S. (1953). *Psychoanalytic theories of personality*: New York, NY, US: McGraw-Hill.

Bolls, P. D., Lang, A., & Potter, R. F. (2001). The Effects of Message Valence and Listener Arousal on Attention, Memory, and Facial Muscular Responses to Radio Advertisements. *Communication Research, 28*, 627-651.

Boyle, G. J. (1989). Re-examination of the major personality-type factors in the Cattell, Comrey and Eysenck scales: Were the factor solutions by Noller et al. optimal? *Personality and Individual Differences, 10*, 1289-1299.

Brackstone, M., & McDonald, M. (1999). Car-following: a historical review. *Transportation Research Part F: Traffic Psychology and Behaviour, 2*, 181-196.

Brandau, H., Daghofer, F., Hofmann, M., & Spitzer, P. (2011). Personality subtypes of young moped drivers, their relationship to risk-taking behavior and involvement in road crashes in an Austrian sample. *Accident Analysis & Prevention, 43*, 1713-1719.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139-159.

Buss, D. M. (1990). Special Issue: Biological Foundations of Personality: Evolution, Behavioral Genetics, and Psychophysiology. *Journal of Personality, 58*.

Byrne, M. D., & Anderson, J. R. (1997). Enhancing ACT-R's perceptual-motor abilites. In *the 19th Annual Conference of the Cognitive Science Society* (pp. 880). Stanford University: Mahwah, NJ: Erlbaum.

Canada, T. A. o. (2006). Performance Measures for Road Networks: A Survey of Canadian Use. In: Transport Canada.

Cary, M., Das, A., Edelman, B., Giotis, I., Heimerl, K., Karlin, A. R., Mathieu, C., & Schwarz, M. (2007). Greedy Bidding Strategies for Keyword Auctions. In *The 8th ACM Conference on Electronic Commerce - EC'07*. San Diego, CA.

Cattell, R. B. (1965). *The scientific analysis of personality*. Baltimore: Penguin Books.

Chandi, K. M. (2005). Sense and Respond Systems. In *The 31st Annual International Conference of the Association of System Performance Professionals*.

Chen, A., Yang, H., Lo, H. K., & Tang, W. H. (2002). Capacity reliability of a road network: an assessment methodology and numerical results. *Transportation Research Part B: Methodological, 36*, 225-252.

Clark, S., & Watling, D. (2005). Modelling network travel time reliability under stochastic demand. *Transportation Research Part B: Methodological, 39,* 119-140.

Clarke, E. H. (1971). Multipart pricing of public goods. *Public Choice, 11,* 17-33.

Clarke, S., & Robertson, I. T. (2005). A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology, 78,* 355-376.

Classen, S., Nichols, A. L., McPeek, R., & Breiner, J. F. (2011). Personality as a predictor of driving performance: An exploratory study. *Transportation Research Part F: Traffic Psychology and Behaviour, 14,* 381-389.

Cloninger, C. R. (1994). Temperament and personality. *Current Opinion in Neurobiology, 4,* 266-273.

CMTT. (2007). Committee for Melbourne Transport Taskforce to the East-West Link Needs Assessment 2007. In: Committee for Melbourne, Victoria, Australia.

CogAff. (2013) (Vol. 2013).

Costa, P., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322-331.

Critchley, H. (2003). Emotion and its disorders: Imaging in clinical neuroscience. *British Medical Bulletin, 65,* 35-47.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Damasio, A. R., Tranel, D., & Damasio, H. C. (1998). Somatic markers and the guidance of behaviour. In J. J.M, O. K. & S. N.L. (Eds.), *Human eomtions: a reader* (pp. 122-135): Blackwell Publishers.

Deery, H. A., & Fildes, B. N. (1999). Young Novice Driver Subtypes: Relationship to High-Risk Behavior, Traffic Accident Record, and Simulator Driving Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 41,* 628-643.

Deffenbacher, J. L., Deffenbacher, D. M., Lynch, R. S., & Richards, T. L. (2003). Anger, aggression, and risky behavior: a comparison of high and low anger drivers. *Behaviour Research and Therapy, 41*, 701-718.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the Structure of Personality: Dopamine, Facilitation of Incentive Motivation, and Extraversion. *Behavioral and Brain Sciences*, *22*, 491-569.

Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science*, *298*, 1191-1194.

Dula, C. S., & Geller, E. S. (2003). Risky, aggressive, or emotional driving: Addressing the need for consistent communication in research. *Journal of Safety Research, 34*, 559-566.

Dyer, M. G. (1991). A society of ideas on cognition: Review of Marvin Minsky's The Society of Mind. *Artificial Intelligence, 48*, 321-334.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion*. Sussex: UK: John Wiley & Sons.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*, 124-129.

Ellison-Potter, P., Bell, P., & Deffenbacher, J. (2001). The Effects of Trait Driving Anger, Anonymity, and Aggressive Stimuli on Aggressive Driving Behavior. *Journal of Applied Social Psychology, 31*, 431-443.

Etzioni, O., & Weld, D. S. (1995). Intelligent agents on the Internet: Fact, fiction, and forecast. *IEEE Expert, 10*, 44-49.

Evans, R., Burke, M., & Dodson, J. (2007). Clothing the Emperor?: Transportation modelling and decision making in Australian cities. In *Urban Research Program*: Griffith University.

Eysenck, H. J. (1983). A biometrical-genetical analysis of impulsive and sensation seeking behavior. In M. Zuckerman (Ed.), *Biological Bases of Sensation Seeking, Impulsivity, and Anxiety* (pp. 1-35). Hillside: NJ: Lawrence Erlbaum Associates.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?--Criteria for a taxonomic paradigm. *Personality and Individual Differences, 12*, 773-790.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, *13*, 667-673.

Fiedler, K., & Bless, H. (2001). The formation of beliefs in the interface of affective and cognitive processes. In N. Frijda, A. Manstead & E. Bem (Eds.), *The Influence of Emotions on Beliefs* (pp. 144-170). Cambridge: Cambridge University Press.

Figueredo, A. J., Sefcek, J. A., Vasquez, G., Brumbach, B. H., King, J. E., & Jacobs, W. J. (2005). Evolutionary personality psychology. In D. M. Buss (Ed.), *Handbook of Evolutionary Psychology*. Hoboken: NJ: Wiley.

Figueredo, A. J., Vásquez, G., Brumbach, B. H., Sefcek, J. A., Kirsner, B. R., & Jacobs, W. J. (2005). The K-factor: Individual differences in life history strategy. *Personality and Individual Differences, 39*, 1349-1360.

Fischer, K., Müller, J. P., & Pischel, M. (1995). Unifying Control In A Layered Agent Architecture. *IJCAI95, AGENT THEORY, ARCHITECTURE AND LANGUAGE WORKSHOP*.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology, 44*, 329-344.

Folke, C., Carpenter, S., Elmqvist, T., Gunderson, L., Holling, C. S., Walker, B., Bengtsson, J., Berkes, F., Colding, J., Danell, K., Falkenmark, M., Gordon, L., Kasperson, R., Kautsky, N., Kinzig, A., Levin, S., Mäler, K. G., Moberg, F., Ohlsson, L., Olsson, P., Ostrom, E., Reid, W., Rockström, J., Savenije, H., & Svedin, U. (2002). Resilience and sustainable development: building adaptive capacity in a world of transformations. In I. C. f. Science (Ed.). Paris, France.

Folke, C., Colding, J., & Berkes, F. (2002). Building resilience for adaptive capacity in social-ecological systems. In F. Berkes, J. Colding & C. Folke (Eds.), *Navigating Social-Ecological Systems: Building Resilience for Complexity and Change*. Cambridge: Cambridge University Press, UK.

Fotherby, T. (2002). Visual Traffic Simulation. In *VISSIM Project - Final Report*. London: Imperial College.

Franklin, S. (1997). AUTONOMOUS AGENTS AS EMBODIED AI. *Cybernetics and Systems, 28*, 499-520.

Franklin, S., & Graesser, A. (1997). Is it an Agent, or just a Program? A Taxonomy for Autonomous Agents. In *Agent Theories, Architectures, and Languages Workshop* (pp. 21-35): Berlin: Springer-Verlag.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology, 62*, 451-482.

Ginsberg, M. (1991). The society of mind: Marvin Minsky. *Artificial Intelligence, 48*, 335-339.

Goldberg, L. R. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology, 59*, 1216-1229.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26-42.

Goodwin, P., & Wright, G. (2009). *Decision Analysis for Management Judgement* (4 ed.): John Wiley and Sons.

Gotlib, I. H., & McCann, C. D. (1984). Construct accessibility and depression: An examination of cognitive and affective factors. *Journal of Personality and Social Psychology, 47*, 427-439.

Government, V. (2010). Victoria population bulletin. In *http://www.dse.vic.gov.au*.

Gregersen, N. P. (1996). Young drivers' overestimation of their own skill--an experiment on the relation between training strategy and skill. *Accident Analysis & Prevention, 28*, 243-250.

Gregersen, N. P., & Bjurulf, P. (1996). Young novice drivers: Towards a model of their accident involvement. *Accident Analysis & Prevention, 28*, 229-241.

Grimm, M., Kroschel, K., Harris, H., Nass, C., Schuller, B., Rigoll, G., & Moosmayr, T. (2007). On the Necessity and Feasibility of Detecting a Driver's Emotional State While Driving. *LNCS, Springer, 4738*, 126–138.

Grimm, M., Kroschel, K., Mower, E., & Narayanan, S. (2007). Primitives-based evaluation and estimation of emotions in speech. *Speech Communication, 49*, 787-800.

Groves, T. (1973). Incentives in teams. *Econometrica, 41*, 617–631.

Haeckel, S. H. (1999). *Adaptive Enterprise: Creating and Leading Sense-and-Respond Organizations*. Cambridge: Harvard Business School Press.

Haidt, J., & Keltner, D. (1999). Culture and Facial Expression: Open-ended Methods Find More Expressions and a Gradient of Recognition. *Cognition & Emotion, 13*, 225-266.

Hermans, H. J. M. (1996). Voicing the self: From information processing to dialogical interchange. *Psychological Bulletin, 119*, 31-50.

Hermans, H. J. M. (2001). The Dialogical Self: Toward a Theory of Personal and Cultural Positioning. *Culture & Psychology, 7,* 243-281.

Hermans, H. J. M. (2002). The Dialogical Self as a Society of Mind: Introduction. *Theory & Psychology, 12*, 147-160.

Ingrand, F. F., Georgeff, M. P., & Rao, A. S. (1992). An architecture for real-time reasoning and system control. *IEEE Expert, 7*, 34-44.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99*, 561-565.

Javela, J. J., Mercadillo, R., & Ramirez, J. M. (2008). Anger and Associated Experiences of Sadness, Fear, Valence, Arousal, and Dominance Evoked by Visual Scenes. *Psychological Reports, 103,* 665-681.

Jenelius, E. (2007). *Approaches to road network vulnerability analysis.* Royal Institute of Technology, Stockholm.

Jenelius, E., Petersen, T., & Mattsson, L.-G. (2006). Importance and exposure in road network vulnerability analysis. *Transportation Research Part A: Policy and Practice, 40*, 537-560.

Jonah, B. A. (1997). Sensation seeking and risky driving: a review and synthesis of the literature. *Accident Analysis & Prevention, 29*, 651-665.

Jovanovic, D., Lipovac, K., Stanojevic, P., & Stanojevic, D. (2011). The effects of personality traits on driving-related anger and aggressive behaviour in traffic among Serbian drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*, 43-53.

Keil, D., & Goldin, D. (2006). Indirect Interaction in Environments for Multi-agent Systems. In D. Weyns, H. Dyke Parunak & F. Michel (Eds.), *Environments for Multi-Agent Systems II* (Vol. 3830, pp. 68-87): Springer Berlin Heidelberg.

Kesting, A., Treiber, M., & Helbing, D. (2007). General Lane-Changing Model MOBIL for Car-Following Models. *Transportation Research Record: Journal of the Transportation Research Board, 1999*, 86-94.

Kleinginna, P. R., & Kleinginna, A. M. (1985). Cognition and affect: A reply to Lazarus and Zajonc. *American Psychologist, 40*, 470-471.

Knoblauch, R. L., Pietrucha, M. T., & Nitzburg, M. (1996). Field Studies of Pedestrian Walking Speed and Start-Up Time. *Transportation Research Record, 1538,* 27-38.

Kubera, Y., Mathieu, P., & Picault, S. (2011). IODA: an interaction-oriented approach for multi-agent based simulations. *Autonomous Agents and Multi-Agent Systems, 23*, 303-343.

Laird, J. E. (2012). The Soar Cognitive Architecture. *AISB Quarterly*, *134*, 1-4.

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

Lajunen, T. (2001). Personality and accident liability: are extraversion, neuroticism and psychoticism related to traffic and occupational fatalities? *Personality and Individual Differences, 31*, 1365-1373.

Lajunen, T., & Parker, D. (2001). Are aggressive people aggressive drivers? A study of the relationship between self-reported general aggressiveness, driver anger and aggressive driving. *Accident Analysis & Prevention, 33*, 243-255.

Lang, A., Dhillon, K., & Dong, Q. (1995). The effects of emotional arousal and valence on television viewers' cognitive capacity and memory. *Journal of Broadcasting & Electronic Media, 39*, 313 - 327.

Laval, J. A., & Daganzo, C. F. (2006). Lane-changing in traffic streams. *Transportation Research Part B: Methodological*, *40*, 251-264.

Lazarus, R. S. (1981). A cognitivist's reply to Zajonc on emotion and cognition. *American Psychologist*, *36*, 222-223.

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. *American Psychologist*, *37*, 1019-1024.

Lazarus, R. S. (1984). On the primacy of cognition. *American Psychologist, 39*, 124-129.

Lerner, J. S., & Keltner, D. (2001). Fear, Anger, and Risk. *Journal of Personality and Social Psychology*, *81*, 146-159.

Lester, D. (2000). National differences in neuroticism and extraversion. *Personality and Individual Differences, 28*, 35-39.

Leu, G., & Namatame, A. (2007). Effectiveness of attack strategies on weighted complex networks. In *The 1st International Conference on Risk Analysis and Crisis Response - RACR07*. Shanghai, China: Atlantis Press.

Lewis, R. L. (2001). Cognitive Theory: SOAR. In J. S. Editors-in-Chief: Neil & B. B. Paul (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 2178-2183). Oxford: Pergamon.

Lisetti, C. L. (2002). Personality, affect and emotion taxonomy for socially intelligent agents. In *FLAIRS 2002*: AAAI.

Luck, M., McBurney, P., Shehory, O., & Willmott, S. (2005). *Agent Technology: Computing as Interaction. A Roadmap for Agent-Based Computing*: the European Coordination Action for Agent-Based Computing (IST-FP6-002006CA).

Luzwick, P. (2000). Situational Awareness and OODA Loops: Coherent Knowledge-based Operations Applied: Last in a series of 4 articles. *Computer Fraud & Security, 2000*, 15-17.

Lynn, R., & Hampson, S. L. (1975). National Differences in Extraversion and Neuroticism. *British Journal of Social and Clinical Psychology, 14*, 223-240.

Lysaker, P. H., & Lysaker, J. T. (2002). Narrative structure in psychosis : Schizophrenia and disruptions in the dialogical self. *Theory and Psychology, 12*, 207–220.

Lyubomirsky, S., Kasri, F., & Zehm, K. (2003). Dysphoric Rumination Impairs Concentration on Academic Tasks. *Cognitive Therapy and Research, 27*, 309-330.

Maes, P. (1991). The agent network architecture (ANA). *SIGART Bull., 2*, 115-120.

Maes, P. (1994). Agents that reduce work and information overload. *Commun. ACM, 37*, 30-40.

Marinier, R. P., Laird, J. E., & Lewis, R. L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research, 10*, 48-69.

Mathew, T. V., & Rao, K. V. K. (2007). Introduction to Transportation Engineering. *NPTEL Program India*.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion, 23*, 209-237.

Mayer, R. E., & Treat, J. R. (1977). Psychological, social and cognitive characteristics of high-risk drivers: A pilot study. *Accident Analysis & Prevention, 9*, 1-8.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 105-126). New York: Kluwer Academic/Plenum.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81-90.

McCrae, R. R., & Terracciano, A. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407-425.

McCrae, R. R., & Terracciano, A. (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. *Journal of Personality and Social Psychology, 88*, 547-561.

McManus, S., Seville, E., Brunsdon, D., & Vargo, J. (2007). Resilience management: a framework for assessing and improving the resilience of organisations. In *Resilient Organisations Research Report* (Vol. 1): Resilient.

McNally, M. G. (2000). The Four-Step Model. In D. A. Hensher & K. J. Button (Eds.), *Handbook of transport modelling* (Vol. 1): Pergamon.

Mershon, B., & Gorsuch, R. L. (1988). Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria? *Journal of Personality and Social Psychology, 55*, 675-680.

Milani, A., & Poggioni, V. (2007). PLANNING IN REACTIVE ENVIRONMENTS. *Computational Intelligence, 23*, 439-463.

Miles, D. E., & Johnson, G. L. (2003). Aggressive driving behaviors: are there psychological and attitudinal predictors? *Transportation Research Part F: Traffic Psychology and Behaviour, 6*, 147-161.

Minsky, M. (1985). *The society of mind*. New York: Simon and Schuster.

Minsky, M. (1991). Society of mind: A response to four reviews. *Artificial Intelligence, 48*, 371-396.

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology, 49*, 229-258.

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*, 723-739.

Ndumu, D. T., & Nwana, H. S. (1997). Research and development challenges for agent-based systems. *Software Engineering. IEE Proceedings*, *144*, 2-10.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newman, M. E. J. (2008). The mathematics of networks. In L. E. Blume & S. N. Durlauf (Eds.), *The New Palgrave Encyclopedia of Economics* (2 ed.). Basingstoke: Palgrave Macmillan.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology, 66*, 574-583.

Ortony, A., & Turner, T. J. (1990). What's basic about basic emotions? *Psychological Review, 97*, 315-331.

Panksepp, J. (2007). Neurologizing the Psychology of Affects: How Appraisal-Based Constructivism and Basic Emotion Theory Can Coexist. *Perspectives on Psychological Science, 2*, 281-296.

Panwai, S., & Dia, H. (2005). Comparative evaluation of microscopic car-following behavior. *IEEE Transactions on Intelligent Transportation Systems, 6*.

Panwai, S., & Dia, H. (2007). Neural Agent Car-Following Models. *Intelligent Transportation Systems, IEEE Transactions on, 8*, 60-70.

Parunak, H. V. D., Brueckner, S., Fleischer, M., & Odell, J. (2003). A preliminary taxonomy of multi-agent interactions. In *AAMAS03*. Melbourne, Australia.

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*, 524-539.

Pêcher, C., Lemercier, C., & Cellier, J.-M. (2009). Emotions drive attention: Effects on driver's behaviour. *Safety Science, 47*, 1254-1259.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nat Rev Neurosci, 9*, 148-158.

Pfeifer, R., & Scheier, C. (1999). *Understanding Intelligence*: MIT Press.

Plomin, R., & Rende, R. (1991). Human Behavioral Genetics. *Annual Review of Psychology, 42*, 161-190.

Rao, A. S., & Georgeff, M. P. (1995). BDI-agents: From Theory to Practice. In *the First International Conference on Multiagent Systems (ICMAS'95)*. San Francisco.

Rapaport, D. (1960). The structure of psychoanalytic theory. *Psychological Issues, 2*, 1-158.

Reeke Jr, G. N. (1991). The society of mind: Marvin Minsky. *Artificial Intelligence, 48*, 341-348.

Revelle, W. (2003). Emotions are to Personality as Weather is to Climate: Analogical reasoning as a tool for scientific investigation. In *The ETS conference on Emotional Intelligence*.

Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *Proceedings of the National Academy of Sciences, 104*, 383-388.

Russell, S., & Norvig, P. (1995). *Artificial Intelligence: a modern approach*. New York: Prentice Hall.

Russello, G., Mostarda, L., & Dulay, N. (2011). A policy-based publish/subscribe middleware for sense-and-react applications. *Journal of Systems and Software, 84*, 638-654.

Sakakibara H., K. Y., and Okada N. (2004). Road Network Robustness for Avoiding Functional Isolation in Disasters. *Journal of Transportation Engineering, 130*, 560.

Salamon, T. (2011). *Design of Agent-Based Models : Developing Computer Simulations for a Better Understanding of Social Processes*: Bruckner Publishing.

Salvucci, D. D. (2006). Modeling Driver Behavior in a Cognitive Architecture. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*, 362-380.

Sarkar, S., Martineau, A., Emami, M., Khatib, M., & Wallace, K. (2000). Aggressive Driving and Road Rage Behaviors on Freeways in San Diego, California: Spatial and Temporal Analyses of Observed and Reported Variations.

Transportation Research Record: Journal of the Transportation Research Board, 1724, 7-13.

Saudino, K. J., & Plomin, R. (1996). Personality and Behavioral Genetics: Where Have We Been and Where Are We Going? *Journal of Research in Personality, 30,* 335-347.

Saulsman, L. M., & Page, A. C. (2004). The five-factor model and personality disorder empirical literature: A meta-analytic review. *Clinical Psychology Review, 23*, 1055-1085.

Scheutz, M., & Logan, B. (2001). Affective vs. Deliberative Agent Control. In *AISB'01 Symposium on Emotion, Cognition and Affective Computing* (pp. 1-10).

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations. *Journal of Cross-Cultural Psychology*, *38,* 173-212.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94,* 168-182.

Shinar, D. (1998). Aggressive driving: the contribution of the drivers and the situation. *Transportation Research Part F: Traffic Psychology and Behaviour, 1,* 137-160.

Silvia, P. J., & Abele, A. E. (2002). Can positive affect induce self-focused attention? Methodological and measurement issues. *Cognition & Emotion*, *16*, 845-853.

Silvia, P. J., Eichstaedt, J., & Phillips, A. G. (2005). Are rumination and reflection types of self-focused attention? *Personality and Individual Differences*, *38,* 871-881.

Sloman, A. (2008). The cognition and affect projetc: architectures, architecture-schemas, and the new science of mind. In. Birmingham, UK: School of Computer Science, University of Birmingham.

Smoliar, S. W. (1991). The society of mind: Marvin Minsky. *Artificial Intelligence, 48*, 349-370.

Stefik, M. J., & Smoliar, S. (1991). Four reviews of The Society of Mind and a response. *Artificial Intelligence*, *48,* 319-320.

Stewart, T. C., & West, R. L. (2006). Deconstructing ACT-R. In *The 7th International Conference on Cognitive Modeling* (pp. 298-303). Trieste, Italy.

Sun, R. (2002). A tutorial on CLARION 5.0. In: Cognitive Science Department, Rensselaer Polytechnic Institute.

Sun, R. (2006). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. In R. Sun (Ed.), *Cognition and Multi-Agent Interaction*. New York: Cambridge University Press.

Sun, R., Peterson, T., & Merrill, E. (1999). A hybrid architecture for situated learning of reactive sequential decision making. *Applied Intelligence, 11*, 109-127.

Teahan, W. J. (2010). *Artificial Intelligence - Agents and Environments*: Ventus Publishing ApS.

Thagard, P. (1993). Societies of minds: Science as distributed computing. *Studies in History and Philosophy of Science Part A, 24*, 49-67.

Thagard, P. (2005). Mind: introduction to cognitive science. *MIT Press, 2nd edition*.

Thagard, P., & Millgram, E. (1995). Inference to the best plan: A coherence theory of decision. In A. Ram & D. B. Leake (Eds.), *Goal-driven learning*. Cambridge: MA: MIT Press.

Tillmann, W. A., & Hobbs, G. E. (1949). The accident-prone automobile driver: A Study of the Psychiatric and Social Background. *Am J Psychiatry, 106*, 321-331.

VanBreda, A. D. (2001). *Resilience theory: A literature review*. Pretoria: South Africa: South African Military Health Service.

Varian, H. R. (2007). Position auctions. *International Journal of Industrial Organization, 25*, 1163-1178.

Velásquez, J. D. (1997). Modeling Emotions and Other Motivations in Synthetic Agents. In *AAAI 97* (pp. 10-15). Providence, Rhode Island: AAAI Press / The MIT Press.

Vickrey, W. (1961). Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance, 16*, 8-37.

VicRoads. (2012). 2010/2011 Performance monitoring information bulletin. In *Network performance monitoring*. Melbourne, Victoria, Australia: VicRoads - Victoria State Road Traffic Authority.

Villagran, d. L. J. C. (2006). Vulnerability: a conceptual and methodological review. *Publications Series of UNU-EHS, 4*.

VPPIA. (2008). Victoria's project prioritisation submission to Infrastructure Australia 2008. In: Department of Transport, State Government of Victoria, Australia.

Watling, D. (2002). A Second Order Stochastic Network Equilibrium Model, I: Theoretical Foundation. *TRANSPORTATION SCIENCE, 36*, 149-166.

Wegner, P. (1997). Why Interaction is More Powerful Than Algorithms. *Communications of the ACM, 40*, 80-91.

Weng, Y., & Wu, T. (2001). Car-following model of vehicular traffic. In *ICII 2001 - International Conferences on Info-tech and Info-net* (Vol. 4, pp. 101-106). Beijing: IEEE Xplore.

Wilson, R. E. (2001). An analysis of Gipps's car-following model of highway traffic. *IMA Journal of Applied Mathematics, 66*, 509–537.

Wooldridge, M. J. (2002). *An Introduction to Multiagent Systems*. Chichester: John Wiley & Sons.

Wooldridge, M. J., & Jennings, N. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review, 10*, 115-152.

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes, 52*, 276-291.

Wu, J., Brackstone, M., & McDonald, M. (2003). The validation of a microscopic simulation model: a methodological case study. *Transportation Research Part C: Emerging Technologies, 11*, 463-479.

Yin, Y., & Ieda, H. (2001). Assessing performance reliability of road networks under nonrecurrent congestion. *Transportation Research Record, 1771*, 148-155.

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist, 39*, 117-123.

Zuckerman, M. (1994). *Behavioural Expressions and Biosocial Bases of Sensation Seeking*. Cambridge: University of Cambridge Press.