Available online at www.sciencedirect.com
ScienceDirect
Cognitive Systems Research 8 (2007) 36–47
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Formal analysis of trace conditioning

Tibor Bosse [a,*], Catholijn M. Jonker [b,1], Sander A. Los [d], Leendert van der Torre [c], Jan Treur [a]

[a] Vrije Universiteit Amsterdam, Department of Artificial Intelligence, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
[b] Radboud Universiteit Nijmegen, Nijmegen Institute for Cognition and Information, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands
[c] University of Luxembourg, Computer Science and Communications (CSC), Faculty of Sciences, Technology and Communication (FSTC), 6, rue Richard Coudenhove, Kalergi, L-1359, Luxembourg
[d] Vrije Universiteit Amsterdam, Department of Cognitive Psychology, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

Received 28 November 2003; received in revised form 12 August 2005; accepted 12 August 2005

Abstract

In the literature classical conditioning is usually described and analysed informally. If formalisation is used, this is often based on mathematical models based on difference or differential equations. This paper explores a formal description and analysis of the process of trace conditioning, based on logical specification and analysis methods of dynamic properties of the process. Specific types of dynamic properties are global dynamic properties, describing properties of the process as a whole, or local dynamic properties, describing properties of basic steps in a conditioning process. If the latter type of properties are specified in an executable format, they provide a temporal declarative specification of a simulation model. By a software environment these local properties can be used to actually perform simulation. Global properties can be checked automatically for simulated or other traces. Using these methods the properties of conditioning processes informally expressed by Los and Van Den Heuvel [Los, S. A., & Van Den Heuvel, C. E. (2001). Intentional and unintentional contributions to non-specific preparation during reaction time foreperiods. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 370–386] have been formalised and verified against a specification of local properties based on Machado's [Machado, A. (1997). Learning the temporal dynamics of behaviour. *Psychological Review*, *104*, 241–265] differential equation model.

Keywords: Trace conditioning; Dynamics; Temporal modelling; Formal analysis; Differential equation

1. Introduction

A common approach to describe dynamics of cognitive processes is by relating sensory, cognitive or behavioural states to previous or subsequent states. For example, this is shown in approaches from what in Philosophy of Mind is called the functionalist perspective, where the functional role of a mental state property is defined by its predecessor and successor states. Also in Dynamical Systems Theory (DST), a relatively new approach to describe the dynamics of cognitive processes, which subsumes connectionist modelling, e.g., Busemeyer and Townsend (1993) and Port and van Gelder (1995), relations of a state with previous and subsequent states are central. van Gelder and Port (1995) briefly explained what a dynamical system is in the following manner. A *system* is a set of changing aspects (or state properties) of the world. A *state* at a given point in time is the way these aspects or state properties are at that time; so a state is characterised by the state properties that hold. The set of all possible states is the *state space*. A *behaviour*

* Corresponding author. Tel.: +31 20 598 7750; fax: +31 20 598 7653. *E-mail address:* tbosse@cs.vu.nl (T. Bosse).
[1] Present address: Delft University of Technology, Department of Mediametics, Man Machine Interaction Group Mekelweg 4, 2628 CD Delft, The Netherlands.

1389-0417/$ - see front matter
doi:10.1016/j.cogsys.2005.08.001

of the system is the change of these state properties over time, or, in other words, a succession or sequence of states within the state space. Such a sequence in the state space can be indexed, for example, by natural numbers (*discrete* case) or real numbers (*continuous* case), and is also called a *trace* or *trajectory*. Given these notions, the notion of *state-determined system*, adopted from (Ashby, 1960) is taken as the basis to describe what a dynamical system is:

> 'A system is state-determined only when its current state always determines a unique future behaviour. (...) the future behaviour cannot depend in any way on whatever states the system might have been in *before* the current state. In other words, past history is irrelevant (or at least, past history only makes a difference insofar as it has left an effect on the current state). (...) the fact that the current state determines future behaviour implies the existence of some *rule of evolution* describing the behaviour of the system as a function of its current state.' (van Gelder & Port, 1995, p. 6).

The assumption of state-based systems, which is fundamental for DST, entails a *local modelling perspective* where states are related to immediate predecessor and successor states. This local modelling perspective, which DST has in common with the functional perspective based on causal relations mentioned earlier, is especially useful for simulation purposes. However, more indirect or *global temporal relations* between states are beyond such a modelling approach.

In more sophisticated cognitive processes, a cognitive state or a behaviour can be better understood in a more global manner, e.g., in the way in which it depends on a longer history of experiences or inputs. In experimental research, examples of such phenomena are usually indicated as inter-trial or adaptive effects. Approaches to model such more sophisticated cognitive processes require means to express a more advanced *temporal complexity* than for the less sophisticated processes, where direct succession relations between states suffice as appropriate means.

Various types of adaptive or learning behaviour are known and have been studied in some depth. For example, for various forms of learned stimulus-response behaviours it has been studied how they are determined by an attained cognitive state of the organism. Still, the question remains how such attained cognitive states themselves depend on the previous history, for example on a particular training session extending over a longer time period. Often insight is obtained by formulating such more complex temporal relationships. However, usually such complex temporal relationships are expressed purely informally, due to the lack of modelling techniques that reach beyond the local perspective.

In order to temporally relate states to states at other points in time in a more global manner, modelling approaches to dynamical systems of a different type have recently been proposed. Within the areas of Computer Science and Artificial Intelligence techniques have been developed to analyse the dynamics of phenomena using logical means. Examples are dynamic and temporal logic, and event and situation calculus; e.g., Eck et al. (2001), Kowalski and Sergot (1986) and Reiter (2001). These logical techniques allow to consider and relate states of a process at different points in time, and in this sense reach beyond the local perspective. The form in which these relations are expressed can cover qualitative aspects, but also quantitative aspects.

This paper addresses temporal aspects of conditioning and illustrates the usefulness of a logical approach for the analysis and formalisation of such processes both at a local and at a more global level. First a local perspective model for temporal conditioning in a high-level executable format is presented. This executable model can be compared to (and was inspired by) Machado (1997)'s differential equation model. Some simulation traces are shown and compared to traces of Machado (1997)'s model.

Next, as part of a non-local perspective analysis, a number of relevant dynamic properties of the conditioning process are identified and formalised. These dynamic properties were obtained by formalising the informally expressed properties to characterise temporal conditioning processes, as put forward by Los and Van Den Heuvel (2001). It has been automatically verified that (under reasonable conditions) these global dynamic properties are satisfied by the simulation traces.

2. Temporal dynamics of conditioning

Below, first some basic terminology used in the area of conditioning is introduced. Next, based on this terminology, Machado (1997)'s differential equation model for conditioning is briefly explained.

2.1. Basic concepts of conditioning

The aim of research into conditioning is to reveal the principles that govern associative learning. To this end, several experimental procedures have been developed. In *classical conditioning*, an organism is presented with an initially neutral conditioned stimulus (e.g., a bell) followed by an unconditioned stimulus (e.g., meat powder) that elicits an innate or learned unconditioned response in the organism (e.g., saliva production for a dog). After acquisition, the organism elicits an adaptive conditioned response (also saliva production in the example) when the conditioned stimulus is presented alone. In *operant conditioning*, the production of a certain operant response that is part to the volitional repertoire of an organism (e.g., bar pressing for a rat) is strengthened after repeated reinforcement (e.g., food presentation) contingent on the operant response.

In their review, Gallistel and Gibbon (2000) argued that these different forms of conditioning have a common foundation in the adaptive timing of the conditioned (or oper-