Studies in Systems, Decision and Control

Volume 394

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control–quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution and exposure which enable both a wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at https://link.springer.com/bookseries/13304

Jan Treur · Laila Van Ments
Editors

Mental Models and Their Dynamics, Adaptation, and Control

A Self-Modeling Network Modeling Approach

Springer

Editors
Jan Treur
Social AI group
Department of Computer Science
Vrije Universiteit Amsterdam, Amsterdam
Noord-Holland, The Netherlands

Laila Van Ments
AutoLeadStar, Jerusalem, Israel

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-3-030-85820-9 ISBN 978-3-030-85821-6 (eBook)
https://doi.org/10.1007/978-3-030-85821-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Formalising human behaviour and mental processes in a formalised format that can be used by computational devices is one of the holy grails in the fields of artificial intelligence and cognitive science. The human brain is so dynamic and complex that it is hard to comprehend that a computer can mimic how it works, an idea that has been the basis for many fantasies, books and movies. This book is meant as a contribution to the journey to this holy grail and to making the gap between human behaviour and computational devices a bit smaller, with a focus in particular on mental processes involving mental models.

Our interests in mental models and their internal simulation and learning were inspired by a number of sources, including literature in areas such as

- Theories and practices in learning psychology, educational science and didactics and in particular the learning and adaptation of mental models addressed in this area
- The concept of as-if body loop for internal mental simulation of body states for emotions
- The concept of simulated perception–action chains for imagination and conscious thought
- The use of theory of mind based on internal mental simulation.

These areas that all concern crucial elements of human mental processes and behaviour prove that it is a challenge to design formalised and computational models for such phenomena involving mental models according to some general architecture. Mental models are often considered as consisting of (networks of) relations between nodes. Such an architecture needs different types of representation for the states involved as nodes in a mental model (internal simulation can take place by their activation) but also for the relations between these states (that trigger the activations of the states). When mental models are used for internal simulation, their states change over time based on these relations. Moreover, these relations adapt when the mental model is learnt, forgotten or revised over time. This suggests that to address mental processes involving mental models, dynamic representations are required both for the states and for the relations. Moreover, on top of that some form of metacognition (metacognitive control) is needed to control these internal simulation and adaptation processes. Together, these are non-trivial and challenging requirements to achieve an appropriate type of formalised computational architecture.

In addition to the above-mentioned areas in the literature, our shared interest in mental models was influenced by a number of case studies addressing specific phenomena that involve mental models, for example, the influence of cognitive metaphors on behaviour and the influence of mental God models on behaviour. Moreover, the idea that a post-traumatic stress disorder is based on a learnt but undesired mental model gave another inspiration. It is in these application cases that some first ideas were developed on how computational models can be designed for mental processes involving mental models. As a first step, using a network-oriented modelling approach, example network models were developed for these three cases, where:

- A mental model was modelled by a subnetwork of an overall mental network model; this subnetwork represents the nodes and relations of the mental model.
- The use of a mental model for internal simulation as part of the overall mental process was modelled as propagation of activation within this subnetwork representing the mental model; outcomes of such forms of internal simulation then can be used as a basis to decide about actions or behaviour.
- Learning or adaptation of the mental model was modelled and implemented separately from the network model, to make the network an adaptive network.
- No metacognitive control of these processes was explicitly modelled.

As a second step, it was found out how adaptive networks can also be modelled in a network-oriented manner by using so-called self-modelling networks (also called reified networks). Self-modelling networks include self-models with nodes representing part of the network's own structure, for example, nodes representing the weights of some of its connections. Using this notion, to introduce new adaptation principles, no new implementations have to be made, as those new adaptation principles can be addressed at the modelling level instead. By using a self-modelling network, for a mental model with its nodes and relations modelled as a (base) subnetwork, the relations of the mental model can explicitly be represented (and adapted) by self-model nodes. This indeed provides the different representations needed for mental processes based on a mental model. Then, in a dedicated software environment, one and the same computational self-modelling network engine can be used in a generic manner to process both the internal simulation of a mental model as propagation of activation within the subnetwork for mental model and the mental model adaptation as propagation of activation within the self-model of the subnetwork for the mental model. Moreover, as the construction of self-modelling networks can easily be iterated to obtain multiple orders of self-models, it was found that also metacognitive control can easily be modelled in this way by a second-order self-model in the self-modeling network.

In recent years, this approach has successfully been applied to model a diversity of specific mental processes based on mental models, as illustrated in many chapters in this volume. In particular, in Parts II (for individual cases) and III (for social cases), each including six chapters. These cover the three examples mentioned above in

Chap. 5 (PTSD as undesired mental model), Chap. 10 (cognitive metaphors as mental models) and Chap. 11 (mental God models). Other chapters in Part II address mental models for learning how a car works and how to drive it (Chap. 3), using metacognitive control to handle multiple mental models (Chap. 4), learning by counterfactual thinking (Chap. 6), complex and adaptive analysis and support tasks (Chap. 7) and self-referencing, self-awareness and self-interpretation within psychotherapy sessions (Chap. 8). Other chapters in Part III address mental models for learner-controlled learning (Chap. 9), attachment theory (Chap. 12), controlled bonding by homophily (Chap. 13) and shared mental models in hospital teams (Chap. 14). Note that there is some overlap between the chapters concerning the general modelling approach used, in order to allow for reading the chapters separately.

In addition, two chapters on philosophical perspectives on mental models are included in Part IV. The first of them is Chap. 15, which proposes context-dependent realisation known from philosophy of mind as a conceptual solution for the issue that different mental models may have different types of neural correlates. The second chapter from a philosophical perspective is Chap. 16, where it is shown how the approach to dynamics and adaptivity used here for mental models can be used to assign emerging informational content to mental models in a way that fits well in the general approaches to dynamics and adaptivity based on temporal factorisation (Treur, 2007), relational specification of mental content (Kim, 1996), and criterial causation (Tse, 2013).

In Part V, some more general background is included on the modelling approach based on self-modelling networks used here: on the supporting software environment (Chap. 17), analysis of self-modelling network behaviour in terms of network structure (Chap. 18), validation of network models by parameter tuning (Chap. 19) and mathematical analysis of the wide applicability of self-modelling networks to model any adaptive dynamical system (Chap. 20).

In Part VI in Chap. 21, an evaluative analysis is presented of what the format of a self-modelling network can offer to the modelling of mental processes involving mental models, covering what already has been achieved and generating options that still are open for further exploration. As an example, while the current book was already in press, for one of those options left open in this book it has been found out by Gülay Canbaloğlu and Jan Treur that the approach introduced here is very useful to obtain in a systematic manner computational models for multilevel organisational learning, an area which lacked such computational models until now. A next book by Gülay and Jan (together with Anna Wiewiora from the area of Management Science) specifically addressing this area based on the modeling approach introduced in the current book will come out later this year or at the beginning of next year.

Finally, we would like to thank Raj Bhalwankar, Gerrit Glas, Peter Roelofsma, Audrey Hermans, Selma Muhammad, and Jan Klein for their contribution to the development of this interesting research area and as authors of one or more chapters

to this book in particular, and Sef van Ments for his help in the final proofreading of the book.

Amsterdam, The Netherlands Jan Treur
Jerusalem, Israel Laila Van Ments
November, 2021

Contents