Representational content and the reciprocal interplay of agent and environment (extended abstract)

Tibor Bosse
Radboud University

Catholijn M. Jonker
Delft University of Technology

Jan Treur
Vrije Universiteit Amsterdam

Representational Content and the Reciprocal Interplay of Agent and Environment (extended abstract)

Tibor Bosse[1], Catholijn M. Jonker[1] and Jan Treur[1,2]

[1] Vrije Universiteit Amsterdam
Department of Artificial Intelligence
De Boelelaan 1081a, 1081 HV Amsterdam
The Netherlands Tel. +31 20 444 77{50,43,63}
{tbosse, jonker, treur}@cs.vu.nl

[2] Utrecht University
Department of Philosophy
Heidelberglaan 8, 3584 CS Utrecht
The Netherlands Tel. +31 30 253 2698
http://www.cs.vu.nl/~{tbosse, jonker, treur}

1. Introduction

Classical approaches to representational content are based on correlations between an agent's internal state properties and external state properties. For example, the presence of a horse in the field is correlated to an internal state property that plays the role of a percept for this horse. One of the critical evaluations of this approach addresses the limitation that internal state properties are to be related to single external states, and cannot be related to processes involving multiple states or events over time. Especially in cases where the agent-environment interaction takes the form of an extensive reciprocal interplay in which both the agent and the environment contribute to the process in a mutual dependency, a classical approach to representational content is insufficient. Some authors even claim that it is a bad idea to aim for a notion of representation in such cases; e.g., [5,7].

As an alternative, in [6], pp. 200-202, the approach to representational content as *relational specification* of an internal state property over time and space is advocated. Using this approach it is possible to relate internal state properties to states at different points in time. Another perspective put forward involving different states over time is Bickhard's *interactivist* approach [1]. In [3] it is shown how a temporal-interactivist approach to representational content of an internal state property can be formalized semantically based on sets of agent-environment past and future interaction trajectories or traces. Moreover, combining the temporal-interactivist perspective with the notion of relational specification, it is shown how to formalise representational content syntactically by characterizing these sets of interaction traces for a given internal state property in terms of relational specifications.

In this paper it is analysed how some non-classical approaches may be used to define representational content in the case of an extensive agent-environment interplay. In particular, for a case study it is discussed how the temporal-interactivist approach and relational specification approach to representational content can be used. These alternative notions involve more complex temporal relationships between internal and external states. Formalisation to define specifications of the representational content more precisely is used as a means to handle this complexity. This formalisation provides dynamic properties that can be (and actually have been) formally checked for given traces of the agent-environment interaction.

2. The Case Study

The case study addressed involves the processes to unlock a front door that sticks. Between the moment that the door is reached and the moment that the door unlocks the following reciprocal interaction takes place:

- the agent puts rotating pressure on the key,
- the door lock generates resistance in the interplay,
- the agent notices the resistance and increases the rotating pressure,
- the door increases the resistance,
- and so on, without any result.

Finally,

- after noticing the impasse the agent changes the strategy by at the same time pulling the door and turning the key, after which the door unlocks.

This example shows different elements. The first part of the process is described in terms of Sun's subconceptual level, whereas the last part of the process is viewed in terms of the conceptual level [7,8].

To model the dynamics of the example, a number of local dynamic properties (in *leads to* format [4]) are considered, describing the basic parts of the process. A subset of them is presented below. In [2] the complete set of properties is given, and it is shown how these properties can be used to generate simulation traces.

LP1 (observation of door). The first local property LP1 expresses that the world state property arriving_at_door leads to an observation of being at the door.

LP2 (observation of resistance). Local property LP2 expresses that the world state property lock_reaction with resistance r leads to an observation of this resistance r.

LP3 (sensory representation of door). Local property LP3 expresses that the observation of being at the door leads to a sensory representation for being at the door.

LP4 (sensory representation of resistance). LP4 expresses that the observation of resistance r of the lock leads to a sensory representation for this resistance.

LP5 (action preparation initiation). LP5 expresses that a sensory representation for being at the door leads to a preparation for the action to turn the key with pressure 1.

LP6 (pressure adaptation). LP6 expresses the following: if turning the key with a certain pressure p did not succeed (since the agent received a resistance that equals p), and the agent has not reached its maximal force (p<mp), and the agent has not learnt anything yet (not c), then it will increase its pressure.

LP7 (birth of learning state). LP7 expresses that, if turning the key with a certain pressure p did not succeed (since the agent received a resistance that equals p), and the agent has reached the limit of its force (p≥mp), then it will learn that should perform a different action (c).

3. Representational Content

In the literature on Philosophy of Mind different types of approaches to representational content of an internal state property have been put forward, for example the correlational, interactivist, relational specification and second-order representation approach; cf. [6], pp. 191-193, 200-202, [1]. These approaches have in common that the occurrence of the internal state property at a specific point in time is related to the occurrence of other state properties, at the same or at different time points. The 'other state properties' can be of three types:

A. *external world state properties*, independent of the agent
B. the agent's sensor state and effector state properties, i.e. the agent's *interaction state properties* (interactivist approach)
C. other *internal state properties* of the agent (higher-order representation)

Furthermore, the type of relationships can be (1) purely functional *one-to-one correspondences*, (e.g., the correlational approach), or (2) they can involve more *complex relationships* with a number of states at different points in time in the past or future, (e.g., the interactivist approach). Moreover, the relationships can be defined (a) at a *semantic* level, by relating semantic structures, or (b) at a *syntactic* level, specifying the relationships in a specific language. So, twelve types of approaches to representational content are distinguished, that can be indicated by codings such as A1a, A1b, and so on. As an example for approach B2a, let us define the representational content of the learning state property c. Thus, we try to specify a temporal relation of internal state property c to sensor and action states in the past. The solution is the following temporal sequence of states: o1, a1(1), o2(1), …, a1(mp), o2(mp). This sequence of states qualifies as a correct representational content for c because it holds iff c holds. Note that by specifying this relationship in a formal language, an approach of type B2b is obtained. The advantage of formalising these relationships is that they can be validated in two ways: (1) by relating them to the local dynamic properties by mathematical proof, and (2) by automatically checking them for simulated traces. In [2], for a large number of the aforementioned approaches to representational content examples are given. All of these examples have been validated.

4. Discussion

In this paper, for some notions of representational content it is explored in a case study of an extensive agent-environment interplay how the temporal complexity can be handled by formalisation. The processes of the case study have been formalised by identifying executable local dynamic properties for the basic dynamics. On the basis of these local properties a simulation model has been made. The formalised specifications of the representational content of the internal state properties have been validated by automatically checking them on the traces generated by the simulation model. Moreover, by mathematical proof it was shown how these specifications are entailed by the basic local properties. This shows that the internal state properties indeed fulfil the representational content specification.

References

[1] Bickhard, M.H., Representational Content in Humans and Machines. *Journal of Experimental and Theoretical Artificial Intelligence*, 5, 1993, pp. 285-333.

[2] Bosse, T., Jonker, C.M., and Treur, J, Representational Content and the Reciprocal Interplay of Agent and Environment. In: R. Sun (ed.), *Proc. of the IJCAI 2003 Workshop on Cognitive Modeling of Agents and Multi-Agent Interactions*, 2003, pp. 19-28.

[3] Jonker, C.M., and Treur, J., A Dynamic Perspective on an Agent's Mental States and Interaction with its Environment. In: C. Castelfranchi and W.L. Johnson (eds.), *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'02.* ACM Press, 2002, pp. 865-872. Extended version: A Temporal-Interactivist Perspective on the Dynamics of Mental States. *Cognitive Systems Research Journal*, vol. 4, 2003, pp. 137-155.

[4] Jonker, C.M., Treur, J., and Wijngaards, W.C.A., A Temporal Modelling Environment for Internally Grounded Beliefs, Desires and Intentions. *Cognitive Systems Research Journal*, vol. 4, 2003, pp. 191-210.

[5] Keijzer, F., Representation in Dynamical and Embodied Cognition. *Cognitive Systems Research Journal*, vol. 3, 2002, pp. 275-288.

[6] Kim, J., *Philosophy of Mind*. Westview Press, 1996.

[7] Sun, R., Symbol grounding: a new look at an old idea. *Philosophical Psychology*, Vol.13, No.2, 2000, pp.149-172.

[8] Sun, R., *Duality of the Mind*. Lawrence Erlbaum Associates, 2002.