Representational Content and the Reciprocal Interplay of Agent and Environment

Tibor Bosse[1], Catholijn M. Jonker[1], and Jan Treur[1,2] **({tbosse, jonker, treur}@cs.vu.nl)**

[1] Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands URL: http://www.cs.vu.nl/~{tbosse, jonker, treur}

[2] Universiteit Utrecht, Department of Philosophy
Heidelberglaan 8, 3584 CS Utrecht, The Netherlands

Abstract

For embodied agents that have an extensive reciprocal interaction with their environment, an internal state does not depend on just one state property of the external world, but is affected both by external aspects of the world and by internal aspects of the agent itself and the way in which these aspects are interwoven during (ongoing) interaction. Therefore, the classical correlational approach to representational content which requires a one-to-one correspondence between an internal state property of an agent and an external world state property is not applicable in such cases. In this paper, a number of alternative notions of representational content are explored with respect to their suitability in such cases. Formal specifications of the representational content of internal state properties have been validated by automatically checking them on traces generated by a simulation model. Moreover, by mathematical proof it was shown how some of these specifications are entailed by the basic local properties.

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 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 claim that it is a bad idea to aim for a notion of representation in such cases; e.g., [Sun, 2000; Keijzer, 2002].

As an alternative, in [Kim, 1996], 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, 1993]'s *interactivist* approach. In [Jonker and Treur, 2002b] 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 will be discussed how the temporal-interactivist approach and relational specification approach to representational content can be used. Moreover, it is shown how to formalise the criteria for representation in terms of dynamic properties that can be formally checked for given (e.g., simulated) traces of the agent-environment interaction.

In this paper, in Section 2 the modelling approach is briefly introduced. Section 3 introduces the case study and the state properties for this case study. In Section 4 the executable local dynamic properties describing basic mechanisms for the case study are presented; simulations on the basis of these local dynamic properties are discussed in Section 5. Section 6 presents dynamic properties for the process as a whole, and for larger parts of the process. In Section 7 the interlevel relations between these nonlocal properties and the local properties are discussed. In Section 8 four different approaches to representational content are explored and formalised. In Section 9 it is shown how these formalisations relate to the local dynamic properties. Section 10 discusses how all these dynamic properties have been checked against the simulation traces. In Section 11 the continuous case is briefly discussed. Section 12 is a discussion.

2 Modelling Approach

To formally specify dynamic properties that express criteria for representational content from a temporal perspective an expressive language is needed. To this end the *Temporal Trace Language* is used as a tool; cf. [Jonker and Treur, 2002a]. In this paper for most of the occurring properties both informal or semi-formal and formal representations are given. The formal representations are based on the Temporal Trace Language (TTL), which is briefly defined as follows.

A *state ontology* is a specification (in order-sorted logic) of a vocabulary, i.e., a signature. A state for ontology Ont is an assignment of truth-values {true, false} to the set At(Ont) of ground atoms expressed in terms of Ont. The *set of all possible states* for state ontology Ont is denoted by STATES(Ont). The set of *state properties* STATPROP(Ont) for state ontology Ont is the set of all propositions over ground atoms from At(Ont). A fixed *time frame* T is assumed which is linearly ordered. A *trace* or *trajectory* $\mathcal{T}$ over a state ontology Ont and time frame T is a mapping $\mathcal{T}$: T → STATES(Ont), i.e., a sequence of states $\mathcal{T}_t$ (t ∈ T) in STATES(Ont). The set of all traces over state ontology Ont is denoted by TRACES(Ont). Depending on the application, the time frame T may be dense (e.g., the real numbers), or discrete (e.g., the set of integers or natural numbers or a finite initial segment of the natural numbers), or any other form, as long as it has a linear ordering. The set of *dynamic properties* DYNPROP(Σ) is the set of temporal statements that can be formulated with respect to traces based on the state ontology Ont in the following manner.

Given a trace $\mathcal{T}$ over state ontology Ont, the input state of the organism (i.e., state of sensors for external world and body) at time point t is denoted by state($\mathcal{T}$, t, input); analogously, state($\mathcal{T}$, t, output), state ($\mathcal{T}$, t, internal) and state ($\mathcal{T}$, t, EW) denote the output state, internal state and external state (of the world, including the physical body) for the organism.

These states can be related to state properties via the formally defined satisfaction relation |=, comparable to the Holds-predicate in the Situation Calculus: state($\mathcal{T}$, t, output) |= p denotes that state property p holds in trace $\mathcal{T}$ at time t in the output state of the organism. Based on these statements, dynamic properties can be formulated in a formal manner in a sorted first-order predicate logic with sorts T for time points, Traces for traces and F for state formulae, using quantifiers over time and the usual first-order logical connectives such as ¬, ∧, ∨, ⇒, ∀, ∃.

To model direct temporal dependencies between two state properties, the simpler 'leads to' format is used. This is an executable format defined as follows. Let α and β be state properties. Then β *follows* α, denoted by α $\twoheadrightarrow_{e, f, g, h}$ β, with time delay interval [e, f] and duration parameters g and h if

$\forall \mathcal{T} \in \mathcal{W}\ \forall t1:$

$\quad \forall t \in [t1 - g, t1) : state(\mathcal{T}, t) \models \alpha \ \Rightarrow$

$\quad \exists d \in [e, f]\ \forall t \in [t1 + d, t1 + d + h) : state(\mathcal{T}, t) \models \beta$

Conversely, the state property β *originates from* state property α, denoted by α $\bullet\!\!\!-\!\!-_{e, f, g, h}$ β, with time delay in [e, f] and duration parameters g and h if

$\forall \mathcal{T} \in \mathcal{W}\ \forall t2:$

$\quad \forall t \in [t2, t2 + h) : state(\mathcal{T}, t) \models \beta \Rightarrow$

$\quad \exists d \in [e, f]\ \forall t \in [t2 - d - g, t2 - d)\ state(\mathcal{T}, t) \models \alpha$

If both α $\twoheadrightarrow_{e,f,g,h}$ β, and α $\bullet\!\!-_{e,f,g,h}$ β hold, then α *leads to* β; this is denoted by: α $\bullet\!\!\twoheadrightarrow_{e,f,g,h}$ β.

The 'leads to' format has shown its value especially when temporal or causal relations in the (continuous) physical world are modelled and simulated in an abstract, non-discrete manner; for example, the intracellular chemistry of *E. coli* (Jonker, Snoep, Treur, Westerhoff and Wijngaards, 2002).

3 The Case Study

In this Section the case study will be introduced and the internal state properties and their dynamics to model this example are presented.

3.1 Introduction of 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, 2000, 2002]'s subconceptual level, whereas the last part of the process is viewed in terms of the conceptual level. For both parts of the process the notion of representational content will be discussed and formalised. The subconceptual part will be used to illustrate both a discrete and a continuous formalisation.

3.2 Internal State properties

To model the example the following internal state properties are used:

s1 sensory representation for being at the door
s2(r) sensory representation for resistance r of the lock
p1(p) preparation for the action to turn the key with rotating pressure p (without pulling the door)
p2 preparation for combined pulling the door and turning the key
c state for having learnt that turning the key should be combined with pulling the door

The interactions between agent and environment are defined by the following sensor and effector states:

o1 observing being at the door
o2(r) observing resistance r
a1(p) action turn the key with rotating pressure p (with out pulling the door)
a2 action turn the key while pulling the door

In addition, the following state properties of the world are used:

arriving_at_door the agent arrives at the door

lock_reaction(r)	the lock reacts with resistance r
door_unlocked	the door is unlocked
d(mr)	resistance threshold mr of the door (indicating that the door will continue to resist until pressure mr or more is used)
max_p(mp)	maximal force on the key that can be exercised by the agent.

4 Local Dynamic Properties

To model the dynamics of the example, the following local properties (in *leads to* format) are considered. They describe the basic parts of the process.

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. Formalisation:

arriving_at_door •↠ o1

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. Formalisation:

lock_reaction(r) •↠ o2(r)

Note that r is a variable here; the specification should be read as a schema for the set of all instances for 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. Formalisation:

o1 •↠ s1

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

o2(r) •↠ s2(r)

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. Formalisation:

s1 •↠ p1(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. Formalisation:

p1(p) and s2(r) and p=r and p<mp and not c •↠ p1(p+1)

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. Formalisation:

p1(p) and s2(r) and p=r and p≥mp •↠ c

LP8 (learning state persistency)
LP8 expresses that the learning state property c persists forever. Formalisation:

c •↠ c

LP9 (alternative action preparation)
LP9 expresses that a sensory representation for resistance r of the lock together with the learning state property lead to a preparation for combined pulling of the door and turning the key. Formalisation:

c and s2(r) •↠ p2

LP10 (action performance)
LP10 expresses that a preparation for the action to turn the key with pressure p (without pulling the door) leads to the actual performance of this action. Formalisation:

p1(p) •↠ a1(p)

LP11 (alternative action performance)
LP11 expresses that a preparation for combined pulling of the door and turning the key leads to the actual performance of this action. Formalisation:

p2 •↠ a2

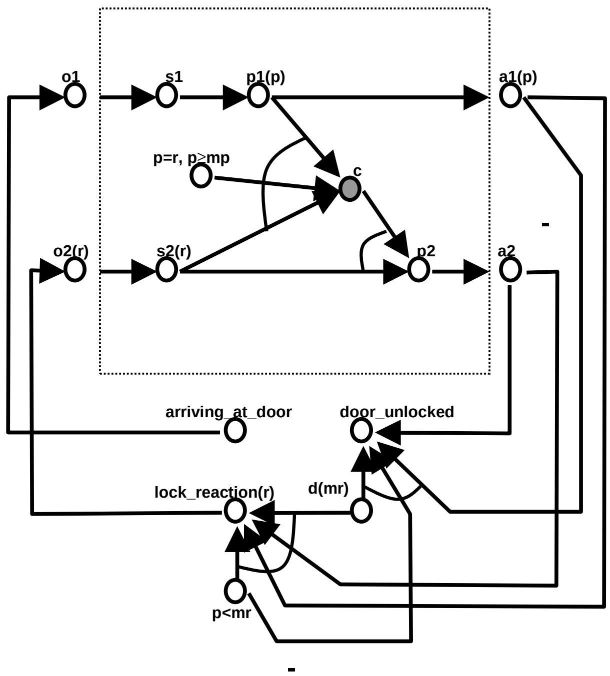

Figure 1 Overview of the simulation model

LP12 (negative effect of action)
LP12 expresses the following property of the world: if the key is turned with a certain pressure p that is smaller than the maximal resistance of the door (p<mr), and the agent is not pulling the door simultaneously, then the lock will react with resistance p. Formalisation:

a1(p) and not a2 and d(mr) and p<mr •→→ lock_reaction(p)

LP13 (positive effect of action)
LP13 expresses the following property of the world: if the key is turned with a certain pressure p that at least equals the maximal resistance of the door (p≥mr), then the door will unlock. Formalisation:

a1(p) and d(mr) and p≥mr •→→ door_unlocked

LP14 (positive effect of alternative action)
LP14 expresses the following property of the world: if the agent turns the key, and simultaneously pulls the door, then the door will unlock. Formalisation:

a2 •→→ door_unlocked

In Figure 1 an overview of these properties is given in a graphical form. To limit complexity, local property LP6 is not depicted.

5 Simulation

A special software environment has been created to enable the simulation of executable models. Based on an input consisting of dynamic properties in *leads to* format, the software environment generates simulation traces. An example of such a trace can be seen in Figure 2. Time is on the horizontal axis, the state properties are on the vertical axis. A dark box on top of the line indicates that the property is true during that time period, and a lighter box below the line indicates that the property is false. This trace is based on all local properties identified above. In property LP6, the values (0,0,1,5) have been chosen for the timing parameters e, f, g, and h. In all other properties, the values (0,0,1,1) have been chosen.

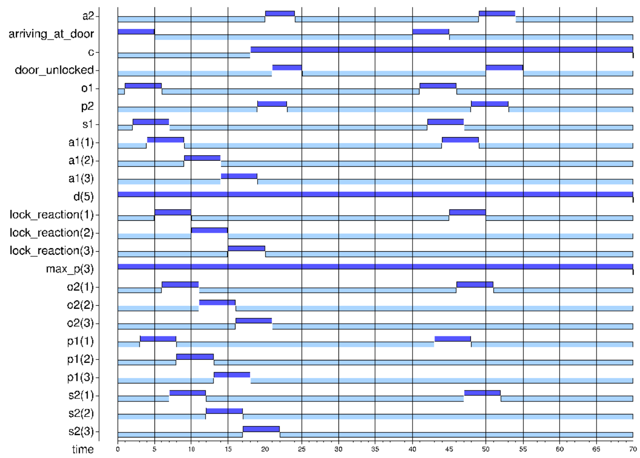

Figure 2 Example simulation trace, where mr=5

As can be seen in Figure 2, the presence of the agent at the door leads to a corresponding observation result (o1), followed by a sensory representation for being at the door. Next, the agent prepares for turning the key (initially with pressure 1), and subsequently performs this action. Since this pressure is insufficient to unlock the door (within this example, the resistant threshold of the door is 5), the door does not open, but a lock reaction (with resistance 1) occurs instead. As a consequence, the agent observes this resistance, and creates a sensory representation of it. At this point, the agent prepares to increase the pressure (see local property LP6), resulting in the action of turning the key with pressure 2. This loop is being activated once more: the agent even tries to turn the key with pressure 3, but then reaches the limit of its force (3 in this example, see LP7) and learns that it should perform a different action. In other words, internal state property c becomes true. Subsequently, the combination of this state property c and state property s2(3) leads to the preparation for an alternative action: combined pulling of the door and turning the key. As a result of this preparation, the action is actually performed and the door is unlocked. However, to show that the agent has indeed learned something, the trace continues for a while. For instance, notice that at time point 40, the agent again finds itself confronted with a locked door. Again, it starts by trying to turn the key with pressure 1. However, when this approach turns out not to work, this time the agent shows adapted behavior. It does not try to increase the pressure, but immediately switches to the alternative action instead. As a reward, it reaches the desired goal of unlocking the door much quicker.

Compared to Figure 2 above, Figure 3 provides a similar trace, but this time the resistant threshold of the door is 2.

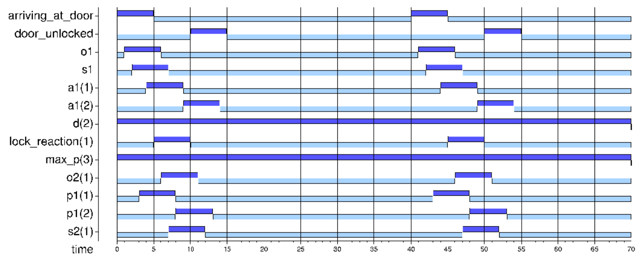

Figure 3 Example simulation trace, where mr=2

By varying this parameter different types of worlds can be simulated, and thus different kinds of behavior can be observed. In this case, the first part of the trace looks the same as in Figure 2, but when the agent tries to turn the key with pressure 2, the door already unlocks. Some time after that, exactly the same situation occurs, and the agent also reacts the same. As a consequence, the agent will never learn the alternative action, simply because there is no need to learn it.

6 Non-local Dynamic Properties

This section presents dynamic properties for larger parts of the process, i.e., at a nonlocal level.

GP1 (door eventually unlocked)
Global property GP1 expresses that eventually the door will be unlocked. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, EW) $\models$ arriving_at_door $\Rightarrow$
$\exists t' \geq t$: state($\mathcal{T}$, t', EW) $\models$ door_unlocked

GP2 (learning occurs)
Global property GP2 expresses that if the maximal resistance of the door is bigger than the maximal rotation force that the agent can exert, then at some point in time learning will occur. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, EW) $\models$ d(mr) $\wedge$
$\forall t$: state($\mathcal{T}$, t, internal) $\models$ max_p(mp) $\wedge$
mr > mp $\Rightarrow$
$\exists t'$ state($\mathcal{T}$, t', internal) $\models$ c

GP2 and the next property GP3 are proved in rather the same way. However, learning as a global process is important enough to represent in its own global property.

GP3 (mr > mp ⇒ door eventually unlocked)
Global property GP3 expresses that if the maximal resistance of the door is bigger than the maximal rotation force that the agent can exert, then at some point in time the door will be unlocked. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, EW) $\models$ d(mr) $\wedge$
$\forall t$: state($\mathcal{T}$, t, internal) $\models$ max_p(mp) $\wedge$
mr > mp $\Rightarrow$
$\exists t'$ state($\mathcal{T}$, t', EW) $\models$ door_unlocked

GP4 (mr ≤ mp ⇒ door eventually unlocked)
Global property GP4 expresses that if the maximal resistance of the door is less than or equal to the maximal rotation force that the agent can exert, then at some point in time the door will be unlocked. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, EW) $\models$ d(mr) $\wedge$
$\forall t$: state($\mathcal{T}$, t, internal) $\models$ max_p(mp) $\wedge$
mr $\leq$ mp $\Rightarrow$
$\exists t'$ state($\mathcal{T}$, t', EW) $\models$ door_unlocked

GP3 and GP4 are formulated separately because their proofs differ. Next a number of intermediate properties are formulated that form a kind of milestones in the process of opening a door and learning.

M1 (at door ⇒ preparation to turn key)
Intermediate property M1 expresses that after the agent stands at the door the agent will prepare for turning the key. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, EW) $\models$ arriving_at_door $\Rightarrow$
$\exists t' > t$: state($\mathcal{T}$, t', internal) $\models$ p1(1)

M2 (lock reaction represented)
Intermediate property M2 expresses that a lock reaction will be represented internally. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, EW) $\models$ lock_reaction(r) $\Rightarrow$
$\exists t' > t$: state($\mathcal{T}$, t', internal) $\models$ s2(r)

M3 (alternative action)
Intermediate property M3 expresses that if lock resistance is internally represented and the agent has learned, then at some later point in time the agent will perform the action a2. Formalisation:
$\forall \mathcal{T} \forall t$: state($\mathcal{T}$, t, internal) $\models$ c $\wedge$ state($\mathcal{T}$, t, internal) $\models$ s2(r) $\Rightarrow$
$\exists t' > t$: state($\mathcal{T}$, t, ouput) $\models$ a2

M4 (increasing rotation pressure)
Intermediate property M4 expresses that under the condition that agent has not learned c yet, the rotation pressure that the agent exerts on the key will always reach the minimum of the maximal resistance of the door and the maximal force that agent can exert. Formalisation:
$\forall \mathcal{T} \forall t, \forall mp, \forall mr, \forall sl$
not state($\mathcal{T}$, t, internal) $\models$ c $\wedge$
state($\mathcal{T}$, t, EW) $\models$ d(mr) $\wedge$
state($\mathcal{T}$, t, internal) $\models$ max_p(mp) $\wedge$
sl = minimum(mr, mp) $\wedge$
state($\mathcal{T}$, t, EW) $\models$ arriving_at_door $\Rightarrow$
$\exists t' > t$: state($\mathcal{T}$, t', internal) $\models$ p1(sl) $\wedge$
$\exists t'' > t'$: state($\mathcal{T}$, t'', output) $\models$ a1(sl)

Finally, a number of additional properties are needed in order to prove the relations between the properties.

A1 (maximal force)
Additional property A1 expresses that the maximal rotation force that the agent can exert on the key is constant. Formalisation:
$\forall \mathcal{T} \exists mp\ \forall t$: state($\mathcal{T}$, t, internal) $\models$ max_p(mp)

A2 (maximal resistance)
Additional property A2 expresses that the maximal resistance that the door can offer is constant. Formalisation:
$\forall \mathcal{T} \exists mr\ \forall t$: state($\mathcal{T}$, t, EW) $\models$ d(mr)

A3 (Closed World Assumption)
The second order property that is commonly known as the Closed World Assumption expresses that at any point in time a state property that is not implied by a specification to be true is false. Let Th be the set of all local properties LP1, through LP14, then the formalisation is:
$\forall P \in$ At(ONT) $\forall \mathcal{T}$ $\forall t$: Th |-/- state($\mathcal{T}$, t) $\models$ P $\Rightarrow$ state($\mathcal{T}$, t) $\models$ not P

7 Interlevel Relations

This section outlines the interlevel connections between dynamic properties at different levels, varying from dynamic properties at the local level of basic parts of the process to dynamic properties at the global level of the overall process. The following interlevel relations between local dynamic properties and nonlocal dynamic properties can be identified.

$$
\begin{array}{ll}
\text{GP3 \& GP4} & \Rightarrow \text{GP1} \\
\text{M2 \& M4 \& LP7 \& LP12} & \Rightarrow \text{GP2} \\
\text{M2 \& M3 \& M4 \& LP7 \& LP14} & \Rightarrow \text{GP3} \\
\text{M4 \& LP13} & \Rightarrow \text{GP4} \\
\text{LP1 \& LP3 \& LP5} & \Rightarrow \text{M1} \\
\text{LP2 \& LP4} & \Rightarrow \text{M2} \\
\text{LP8 \& LP9 \& LP11} & \Rightarrow \text{M3} \\
\text{M1 \& M2 \& LP6 \& LP10 \& LP12 \& A1 \& A2 \& A3} & \Rightarrow \text{M4}
\end{array}
$$

The proofs of M1, M2, M3, and GP1 are rather straightforward and left out. A proof sketch of the other properties is provided.

Property M4 can be proved by induction. The induction hypothesis is

$$
\begin{aligned}
&\forall \mathcal{T} \forall t: \text{state}(\mathcal{T}, t, \text{output}) \models a1(p) \wedge p < sl \Rightarrow \\
&\quad \exists t1 > t, \exists t2 > t1: \\
&\qquad \text{state}(\mathcal{T}, t1, \text{internal}) \models p1(p+1) \wedge \\
&\qquad \text{state}(\mathcal{T}, t2, \text{output}) \models a1(p+1)
\end{aligned}
$$

The induction base is given by properties M1 and LP10, providing p1(1), and a1(1). The induction step is proved along the following lines. Because of A3 and "not c", also "not a2" holds at all times during which "not c" holds. Since $p < sl$ and sl is the minimum of mp and mr, $p < mr$. All conditions of LP12 hold, therefore, some time after a1(p) holds, lock_reaction(p) will hold. Applying property M2 tells us that again some time later s2(p) will hold. Since p is also less than mp now LP6 can be applied giving us that again some time later (call this time point t1) p1(p+1) will hold. By applying LP10, we now have that some time later again, say at time t2, a1(p+1) will hold. This concludes the proof by induction of M4.

Property GP2 can be proved as follows. Since $mr > mp$, $sl = mp$. Applying M4 gives us

$\forall \mathcal{T} \exists t'$: state(γ, t', output) $\models$ a1(mp).

By application of LP12, we get some time later lock_reaction(mp), application of M2 gives us, some time later again, s2(mp). Finally, application of LP7 provides us with the learned c.

The proof of Property GP3 follows the following subsequent time points of interest:

Application of M4 gives a time point t1 such that p1(mp) holds, application of M2 give a time t2 such that s2(mp) holds, application of LP7 gives a time t3 such that c holds, application of M3 gives a time t4 such that a2 holds, application of LP14 gives a time t5 such that door_unlocked holds.

The proof of property GP4 is rather short, by application of M4 at a certain time t1 a1(mr) will hold, by application of LP13 a later time t2 exist at which door_unlocked holds.

All proofs can be worked out in more details by using the timing parameters of the local properties involved.

8 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. [Kim, 1996], pp. 191-193, 200-202, [Bickhard, 1993]. These approaches to representational content 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 contents are distinguished, that can be indicated by codings such as A1a, A1b, and so on. Below, for a number of such approaches examples are given.

8.1 Correlational Approach

According to the Correlational approach, the representational content of a certain internal state is given by a one-to-one correlation to another (in principle external) state property: type A1a. Such an external state property may exist backward as well as forward in time. Hence, for the current example, in order to define the representational content of an internal state property, one should try if this can be related to a world state property that either existed in the past or will exist in the future. For example, the representational content for internal state property s1 can be defined as world state property arriving_at_door, by looking backward in time. Intuitively, this is a correct definition, since for all possible situations where the agent has s1, it was indeed physically present at the door, and conversely. Likewise, the representational content for internal state property p2 can be defined as action property a2, by looking forward in time, or, rather, as world state property door_unlocked. However, for many other internal state properties the representational content cannot be defined adequately according to the correlational approach. In these cases, reference should not be made to one single state in the past or in the future, but to a temporal sequence of inputs or output state properties, which is not considered to adequately fit in the correlational approach. An overview for the content of all internal state properties according to the correlational approach (if any), is given in Table 1. These relationships can easily be specified in the language TTL, thus an approach of type A1b is obtained.

Internal state property	Content (backward)	Content (forward)
s1	arriving_at_door	lock_reaction(1)
s2(r)	lock_reaction(r)	*impossible*
p1(1)	arriving_at_door	lock_reaction(1)
p1(2)	*impossible*	lock_reaction(2)
p2	*impossible*	door_unlocked
c	*impossible*	*impossible*

Table 1 Correlational approach

8.2 Temporal-Interactivist Approach

The Temporal-Interactivist approach [Bickhard, 1993; Jonker and Treur, 2002] relates the occurrence of internal state properties to sets of past and future interaction traces: type B. This can be done in the form of functional one-to-one correspondences (type B1), or by involving more complex relationships over time (type B2). Moreover, these relationships can be defined at a semantic level (type B1a or B2a), or at a syntactic level (type B1b or B2b). In this paper the focus is on the B2 type, which is the more advanced case. First the semantic variant B2a is addressed, next the syntactic variant B2b. As an example, consider the internal state property c. The representational content of c is defined in a semantic manner by the pair of sets of past interaction traces and future interaction traces (here InteractionOnt denotes the input and output state ontology and IntOnt the internal state ontology):

PITRACES(c) =

$$\{ \mathcal{T}_{\leq t}^{\text{InteractionOnt}} \mid t \in T, \text{state}(\mathcal{T}, t, \text{IntOnt}) \models c \}$$

FITRACES(c) =

$$\{ \mathcal{T}_{\geq t}^{\text{InteractionOnt}} \mid t \in T, \text{state}(\mathcal{T}, t, \text{IntOnt}) \models c \}$$

Here the first set, PITRACES(c), contains all past interaction traces for which sequence of time points exists such that at these time points first o1 occurs, next a1(1), next o2(1), next a1(2), next o2(2), next a1(3), and next o2(3). For this example, a learning phase of 3 trials has been chosen. The second set, FITRACES(c), contains all future interaction traces for which no o2(r) occurs, or o2(r) occurs and after this a2 occurs. It is precisely these descriptions of the sets of interaction traces that are formalised in the syntactic variant, in order to obtain a description of type B2b.

8.3 Relational Specification Approach

The Relational Specification approach to representational content is based on a specification of how the occurrence of an internal state property relates to properties of states distant in space and time; cf. [Kim, 1996], pp. 200-202. This is a quite general concept; however in this paper it is used in conjunction with the temporal-interactivist approach. Thus, according to the Temporal-Interactivist Relational Specification approach, the representational content of a certain internal state can be defined by specifying a temporal relation of the internal state property to sensor and action states in the past and future. An overview for the content of all internal state properties according to the temporal relational specification approach is given, in an informal notation, in Table 2. Note that these relationships are defined at a semantic level, and are thus (again) of type B2a. Different interaction state properties, separated by commas, should be read as the temporal sequence of these states. For this example, a learning phase of 3 trials has been chosen.

Internal State Property	Content (backward)	Content (forward)
s1	o1	a1(1)
s2(r)	o2(r)	if c (defined by o1, …, o2(3)), then a2
p1(1)	o1	a1(1)
p1(2)	o1, a1(1), o2(1)	a1(2)
p1(3)	o1, a1(1), o2(1), a1(2), o2(2)	a1(3)
p2	o1, a1(1), o2(1), a1(2), o2(2), a1(3), o2(3)	a2
c	o1, a1(1), o2(1), a1(2), o2(2), a1(3), o2(3)	if o2(r), then a2

Table 2 Temporal-interactivist approach (semantic level)

Table 3 and 4 describe the same information as Table 2, but this time characterised by formulae in a specific language, TTL in out case. Thus, the relationships are defined at a syntactic level, and are of type B2b. The following abstractions are used:

is_followed_by($\mathcal{T}$, X, I1, Y, I2) ≡
∀t1: state($\mathcal{T}$, t1, I1) ⊨ X ⇒
  ∃t2 ≥ t1: state($\mathcal{T}$, t2, I2) ⊨ Y

This expresses that X is always followed by Y.

is_preceded by($\mathcal{T}$, Y, I1, X, I2) ≡
∀t1: state($\mathcal{T}$, t2, I1) ⊨ Y ⇒
  ∃t1 ≤ t2: state($\mathcal{T}$, t1, I2) ⊨ X

This expresses that Y is always preceded by X.

These abstractions can be used like is_preceded_by($\mathcal{T}$, s1, internal, o1, input), is_followed_by($\mathcal{T}$, o2(1), input, s2(1), internal), et cetera. The next abstraction describes that the interplay between agent and environment in which the agent increases pressure and the environment increases resistance is performed up to a certain level of pressure.

interplay_up_to($\mathcal{T}$, t1, t2, 1) ≡ t1≤ t2 &
  state($\mathcal{T}$, t1, output) ⊨ a1(1) & state($\mathcal{T}$, t2, input) ⊨ o2(1)

interplay_up_to($\mathcal{T}$, t1, t4, 2) ≡
∃t2, t3 [t1 ≤ t2 ≤ t3 ≤ t4]
interplay_up_to($\mathcal{T}$, t1, t2, 1) &
state($\mathcal{T}$, t3, output) ⊨ a1(2) & state($\mathcal{T}$, t4, input) ⊨ o2(2)

interplay_up_to($\mathcal{T}$, t1, t6, 3) ≡
∃t4, t5 [t1 ≤ t4 ≤ t5 ≤ t6]
interplay_up_to($\mathcal{T}$, t1, t4, 2) &
state($\mathcal{T}$, t5, output) ⊨ a1(3) & state($\mathcal{T}$, t6, input) ⊨ o2(3)

I.s.p.	Content (backward)
s1	is_followed_by($\mathcal{T}$, o1, input, s1, internal) & is_preceded_by($\mathcal{T}$, s1, internal, o1, input)
s2(r)	is_followed_by($\mathcal{T}$, o2(r), input, s2(r), internal) & is_preceded by($\mathcal{T}$, s2(r), internal, o2(r), input)
p1(1)	is_followed_by($\mathcal{T}$, o1, input, p1(1), internal) & is_preceded by($\mathcal{T}$, p1(1), internal, o1, input)
p1(2)	∀t1,t2,t3 [t1≤t2≤t3 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t3,1) & not [∃t11,t12,t17 [t11≤t12≤t17≤t3 & state($\mathcal{T}$, t11, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t12, t17,3)]] ⇒ ∃t4 ≥ t3 state($\mathcal{T}$, t4, internal) ⊨ p1(2)] & ∀t4 [state($\mathcal{T}$, t4, internal) ⊨ p1(2) ⇒ ∃t1,t2,t3 t1≤t2≤t3≤t4 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t3,1)]
p1(3)	∀t1,t2,t5 [t1≤t2≤t5 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t5, 2) ⇒ ∃t6 ≥ t5 state($\mathcal{T}$, t6, internal) ⊨ p1(3)] & ∀t6 [state($\mathcal{T}$, t6, internal) ⊨ p1(3) ⇒ ∃t1,t2,t5 t1≤t2≤t5≤t6 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t5,2)]
p2	∀t1,t2,t7 [t1≤t2≤t7 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t7,3) ⇒ ∃t8 ≥ t7 state($\mathcal{T}$, t8, internal) ⊨ p2] & ∀t8 [state($\mathcal{T}$, t8, internal) |== p2 ⇒ ∃t1,t2,t7 t1≤t2≤t7≤t8 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t7,3)]
c	∀t1,t2,t7 [t1≤t2≤t7 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t7,3) ⇒ ∃t8 ≥ t7 state($\mathcal{T}$, t8, internal) ⊨ c] & ∀t8 [state($\mathcal{T}$, t8, internal) ⊨ c ⇒ ∃t1,t2,t7 t1≤t2≤t7≤t8 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t7,3)]

Table 3 Temporal-interactivist relational specification approach (syntactic level, backward)

I.s.p.	Content (forward)
s1	is_followed_by($\mathcal{T}$, s1, internal, a1(1), output) & is_preceded by($\mathcal{T}$, a1(1), output, s1, internal)
s2(r)	∀t0 [state($\mathcal{T}$, t0, internal) ⊨ s2(3) ⇒ ∀t1,t2,t3,t7 [t1≤t2≤t7 & t7 ≥ t0 & state($\mathcal{T}$, t1, input) ⊨ o1 & interplay_up_to($\mathcal{T}$, t2, t7,3) ⇒ ∃t8 ≥ t7 state($\mathcal{T}$, t8, output) ⊨ a2]]
p1(1)	is_followed_by($\mathcal{T}$, p1(1), internal, a1(1), output) & is_preceded by($\mathcal{T}$, a1(1), output, p1(1), internal)
p1(2)	is_followed_by($\mathcal{T}$, p1(2), internal, a1(2), output) & is_preceded by($\mathcal{T}$, a1(2), output, p1(2), internal)
p1(3)	is_followed_by($\mathcal{T}$, p1(3), internal, a1(3), output) & is_preceded by($\mathcal{T}$, a1(3), output, p1(3), internal)
p2	is_followed_by($\mathcal{T}$, p2, internal, a2, output) & is_preceded by($\mathcal{T}$, a2, output, p2, internal)
c	∀t1 [state($\mathcal{T}$, t1, internal) ⊨ c ⇒ ∀t2 ≥ t1 [state($\mathcal{T}$, t2, input) ⊨ o2(r) ⇒ ∃t3 ≥ t2 state($\mathcal{T}$, t3, output) ⊨ a2]]

Table 4 Temporal-interactivist relational specification approach (syntactic level, forward)

8.4 Second-Order Representation

In approaches to representational content of type C, internal state properties are related to other internal state properties. For example, in [Sun, 2000, 2002]'s dual approach to cognition, conceptual level state properties are related to subconceptual level state properties:

> *On this view, high-level conceptual, symbolic representation is rooted, or grounded, in low-level behavior (comportment) from which it obtains its meanings and for which it provides support and explanations. The rootedness/groundedness is guaranteed by the way high-level representation is produced: It is, in the main, extracted out of low-level behavioral structures.*
> [Sun, 2000]

Two possibilities arise: either the other internal state properties are not considered to be representational (this seems to be Sun's position), or they are themselves considered representations of something else. In the latter case, which is explored here, the conceptual level state properties become second-order representations: representations of representations. In the main example of this paper, the internal state property c can be considered to be at the conceptual level, whereas the other, s and p properties are considered subconceptual. Then, in the spirit of [Sun, 2000], the representational content of c can be defined in terms of the other internal state properties as follows.

Backward:
c will occur if in the past once s1 occurred, then p1(1), then s2(1), then p1(2), then s2(2), then p1(3), then s2(3), and conversely. Formally:

∀t1,t2,t3,t4,t5,t6,t7 [t1≤t2≤t3≤t4≤t5≤t6≤t7
& state($\mathcal{T}$, t1, internal) ⊨ s1
& state($\mathcal{T}$, t2, internal) ⊨ p1(1) & state($\mathcal{T}$, t3, internal) ⊨ s2(1)
& state($\mathcal{T}$, t4, internal) ⊨ p1(2) & state($\mathcal{T}$, t5, internal) ⊨ s2(2)
& state($\mathcal{T}$, t6, internal) ⊨ p1(3) & state($\mathcal{T}$, t7, internal) ⊨ s2(3)
⇒ ∃t8 ≥ t7 state($\mathcal{T}$, t8, internal) ⊨ c]
&

$$\forall t8\ [\ \text{state}(\mathcal{T}, t8, \text{internal}) \models c \Rightarrow$$
$$\exists t1,t2,t3,t4,t5,t6,t7\ \ t1 \leq t2 \leq t3 \leq t4 \leq t5 \leq t6 \leq t7 \leq t8$$
$$\&\ \text{state}(\mathcal{T}, t1, \text{internal}) \models s1$$
$$\&\ \text{state}(\mathcal{T}, t2, \text{internal}) \models p1(1)\ \&\ \text{state}(\mathcal{T}, t3, \text{internal}) \models s2(1)$$
$$\&\ \text{state}(\mathcal{T}, t4, \text{internal}) \models p1(2)\ \&\ \text{state}(\mathcal{T}, t5, \text{internal}) \models s2(2)$$
$$\&\ \text{state}(\mathcal{T}, t6, \text{internal}) \models p1(3)\ \&\ \text{state}(\mathcal{T}, t7, \text{internal}) \models s2(3)\ ]$$

Forward:
If c occurs, then in the future, if s2(r) occurs, then p2 will occur. Formally:

$$\forall t1\ [\ \text{state}(\mathcal{T}, t1, \text{internal}) \models c \Rightarrow$$
$$\forall t2 \geq t1\ [\ \text{state}(\mathcal{T}, t2, \text{internal}) \models s2(r) \Rightarrow$$
$$\exists t3 \geq t2\ \text{state}(\mathcal{T}, t3, \text{internal}) \models p2\ ]\ ]$$

9 Relations between Relational Specifications and Local Properties

The specifications of representational content have been validated in two ways: (1) by relating them to the local dynamic properties by mathematical proof, and (2) by automatically checking them for the simulated traces (see Section 10). The former is briefly shown below.

Correlational approach
According to Table 1 state property p1 represents the backward content arriving_at_door. This choice can be proven by the validity of property M1 and its inverse:
$$\forall\ \mathcal{T} \forall t:\ \ \text{state}(\mathcal{T}, t, \text{internal}) \models p1(1) \Rightarrow$$
$$\exists t' \leq t:\ \text{state}(\mathcal{T}, t', \text{EW}) \models \text{arriving\_at\_door}$$
This reverse property can be proven from A3 and all local properties, explaining that the only way to obtain p1(1) is through M1.

Temporal-interactivist relational specification approach
Consider the formula that presents the backward representational content for internal state property c in Table 3. Consider first the direction from observations to c. Given o1, o2(1), o2(2), and o2(3) at the different subsequent time points the proof obligation is c. Given o1, by applying (in this order) LP3, LP5 we obtain p1(1) which we need to derive from the given o2(1) using LP4, s2(1) and by application of LP6 on p1(1) and s2(1) we obtain p1(2). Given o2(2), by application of LP4 we obtain s2(2) and on the basis of p1(2) LP6 is again applicable resolving into p1(3). Given o2(3), apply LP4 to obtain s2(3), and using p1(3) LP7 is applicable and c obtained. These dependencies are graphically represented in Figure 4. The reverse direction again depends on property A3 and all local properties.

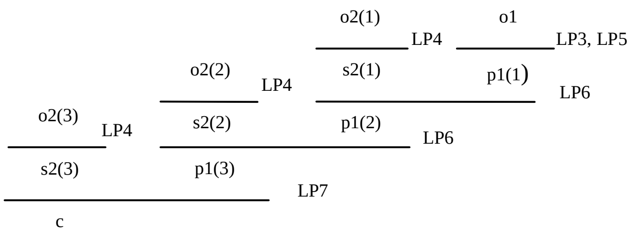

Figure 4 Proof Tree

10 Checking Properties

In addition to the software described in Section 5, other software has been developed that takes traces and formally specified properties as input and checks whether a property holds for a trace. Using automatic checks of this kind, many of the properties presented in this paper have been checked against the traces depicted in Figure 2 and 3. In particular, the global properties GP1, GP2, GP3, and GP4, and the intermediate properties M1, M2, M3, and M4 have been checked, and all turned out to hold for the given traces. Furthermore, all properties for representational content denoted in Table 3 and 4 have been checked. Success of these checks would validate our choice for the representational content (according to the relational specification approach) of the internal state properties s1, s2(r), p1(1), p1(2), p1(3), p2, and c. Initially, one of these checks did not succeed. It turned out that the backward representational content defined for p1(2) was not correctly chosen. At that time, this content was defined as follows:

$$\forall t1,t2,t3\ [\ t1 \leq t2 \leq t3\ \&\ \text{state}(\mathcal{T}, t1, \text{input}) \models o1\ \&$$
$$\text{state}(\mathcal{T}, t2, \text{output}) \models a1(1)\ \&\ \text{state}(\mathcal{T}, t3, \text{input}) \models o2(1)$$
$$\Rightarrow \exists t4 \geq t3\ \text{state}(\mathcal{T}, t4, \text{internal}) \models p1(2)\ ]$$
$$\&\ \forall t4\ [\ \text{state}(\mathcal{T} t4, \text{internal}) \models p1(2) \Rightarrow$$
$$\exists t1,t2,t3\ \ t1 \leq t2 \leq t3 \leq t4\ \&\ \text{state}(\mathcal{T}, t1, \text{input}) \models o1\ \&$$
$$\text{state}(\mathcal{T}, t2, \text{output})\ |{=}{=}\ a1(1)\ \&\ \text{state}(\mathcal{T}, t3, \text{input}) \models o2(1)\ ]$$

According to, the above notation, the sequential occurrence of the state properties o1, o1(1), and o2(1) always implies that state property p1(2) will occur. However, a close examination of Figure 2 reveals that this is not always the case. Whenever the agent has learned, it will not increase its pressure on the key anymore. As a result, the extra condition not c had to be added to the representational content. All the other checks concerning the properties of Table 3 and 4 did succeed immediately.

11 Continuous Dynamics

The agent-environment interaction limited to turning the key without pulling the door can be described in a continuous manner using the following equations E:

$$dp(t)/dt = \alpha\ r(t)\ (P - p(t)) \qquad \text{(agent)}$$
$$r(t) = p(t) \qquad \text{(environment)}$$

Here it is assumed that the pressure p at time t is increased by the agent proportionally to the observed resistance r at time t, until a maximal pressure P, and the resistance by the environment is simply equal to the pressure (standard law of action = reaction). Possible trajectories are real-valued functions $p : T \rightarrow R$, $r : T \rightarrow R$ that satisfy the above equations (here R denotes the real numbers).

Based on this continuous formalisation the question can be addressed what for a certain pressure value v the representational content of p1(v) is. If a temporal-interactivist relational perspective is taken, the occurrence of the internal state p1(v) at a time point t relates to the set of past trajectories that is described by

$\{(p, r) \mid p: T_{\leq t} \rightarrow R,\ r : T_{\leq t} \rightarrow R$ satisfy E and $p(t) = v\}$
and at the same time the set of future trajectories described by
$\{(p, r) \mid p: T_{\geq t} \rightarrow R,\ r : T_{\geq t} \rightarrow R$ satisfy E and $p(t) = v\}$

12 Discussion

The classical correlational approach to representational contents requires a one-to-one correspondence between an internal state property of an agent and an external world state property. For embodied agents that have an extensive reciprocal interaction with their environment, this classical correlational approach does not suffice. In particular, an internal state in such an agent does not depend on just one state property of the external world, but is affected both by external aspects of the world and by internal aspects of the agent itself and the way in which these aspects are interwoven during the (ongoing) interaction process.

Given this problem, it is under debate among several authors whether adequate alternative notions of representational content exist for such an embodied agent's internal states. Some authors claim that for at least part of the internal states it makes no sense to consider them as conceptual or as having representational content; e.g., [Clark, 1997; Sun, 2000; Keijzer, 2002]. Other authors claim that some notions of representational content can be defined, but these strongly deviate from the classical correlational approach; e.g., [Bickhard, 1993; Kim, 1996; Jonker and Treur, 2002].

In this paper, for a number of notions of representational content it was explored in a case study how they work out, and, especially, how they can be formalised. 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 specifications of the representational content of the internal state properties have been validated for the discrete case 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.

As a final remark, notice that the forward representational content of both property s2(r) and c (in Table 4) is only defined in a single direction. For instance, in the case of c, one could be tempted to include the following property:

$\forall t2$ [[state($\mathcal{T}$, t2, internal) $\models$ o2(r) $\Rightarrow$
  $\exists t3 \geq t2$ state($\mathcal{T}$, t3, EW) $\models$ a2] $\Rightarrow$
    $\exists t1 \leq t2$ state($\mathcal{T}$, t1, internal) $\models$ c]

This property has the pattern [A=>B]=>C. However, if A does not occur then such a property would force C to be true, irrespective of whether B occurs. This would not lead to an appropriate situation, since in traces where never resistance of the key is observed, there is no need to have internal state property c. This dynamic property could be replaced, however, by a more complicated dynamic property involving quantification over traces, as shown in [Jonker and Treur, 2002b].

References

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

[Clark, 1997] Clark, A., (1997). *Being There: Putting Brain, Body and World Together Again*. MIT Press, 1997.

[Jonker et al., 2002] Jonker, C.M., Snoep, J.L., Treur, J., Westerhoff, H.V., and Wijngaards, W.C.A., (2002). BDI-Modelling of Intracellular Dynamics. In: A.B. Williams and K. Decker (eds.), *Proc. of the First International Workshop on Bioinformatics and Multi-Agent Systems, BIXMAS'02*, 2002, pp. 15-23.

[Jonker and Treur, 2002a] Jonker, C.M. and Treur, J., Compositional Verification of Multi-Agent Systems: a Formal Analysis of Pro-activeness and Reactiveness. *International Journal of Cooperative Information Systems*, vol. 11, 2002, pp. 51-92.

[Jonker and Treur, 2002b] Jonker, C.M., and Treur, J., (2002). 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 to appear as: A Temporal-Interactivist Perspective on the Dynamics of Mental States. *Cognitive Systems Research Journal*. To appear, 2003.

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

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

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

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