Simulation and Representation of Body, Emotion, and Core Consciousness

Tibor Bosse[*] Catholijn M. Jonker[†] Jan Treur[*]

[*]Vrije Universiteit Amsterdam
Department of Artificial Intelligence
De Boelelaan 1081a
1081 HV Amsterdam
The Netherlands
{tbosse, treur}@cs.vu.nl

[†]Radboud Universiteit Nijmegen
Nijmegen Institute for Cognition and Information
Montessorilaan 3
6525 HR Nijmegen
The Netherlands
C.Jonker@nici.ru.nl

Abstract

This paper contributes an analysis and formalisation of Damasio's theory on core consciousness. Three important concepts in this theory are "emotion", "feeling", and "feeling a feeling" (or core consciousness). In particular, a simulation model is described of the neural dynamics leading via emotion and feeling to core consciousness, and dynamic properties are formally specified that hold for these dynamics. These properties have been automatically checked for the simulation traces. Moreover, a formal analysis is made and verified of relevant notions of representation.

1 Introduction

In (Damasio, 2000) the neurologist Antonio Damasio puts forward his theory of consciousness. He describes his theory in an informal manner, and supports it by a vast amount of evidence from neurological practice. More experimental work supporting his theory is reported in (Damasio et al., 2000; Parvizi and Damasio, 2001). Damasio's theory is described on the one hand in terms of the occurrence of certain neural states (or neural patterns), and temporal or causal relationships between them. Formalisation of these relationships requires a modelling format that is able to express direct temporal or causal dependencies. On the other hand Damasio gives interpretations of most of these neural states as representations, for example as 'sensory representation', or 'second-order representation'. This requires an analysis of what it means that a neural state is a representation for something. This paper focuses on Damasio's notions of 'emotion', 'feeling', and 'core consciousness' or 'feeling a feeling'. In (Damasio, 2000), Damasio describes an *emotion as neural object* (or *internal emotional state*) as an (unconscious) neural reaction to a certain stimulus, realized by a complex ensemble of neural activations in the brain. As the neural activations involved often are preparations for (body) actions, as a consequence of an internal emotional state, the body will be modified into an *externally observable emotional state*. Next, a *feeling* is described as the (still unconscious) sensing of this body state. Finally, *core consciousness* or *feeling* a *feeling* is what emerges when the organism detects that its representation of its own body state (the *proto-self*) has been changed by the occurrence of the stimulus: it becomes (consciously) aware of the feeling.

This paper aims at formalisations and simulation models for these three notions. In addition, the notion of representation used by Damasio is formally analysed against different approaches to representational content from the literature on the Philosophy of Mind. It is shown that the classical causal/correlational approach to representational content, e.g., (Kim, 1996), pp. 191-193, is inappropriate to describe the notion of representation for core consciousness used by Damasio, as this notion essentially involves more complex temporal relationships describing histories of the organism's interaction with the world. An alternative approach is shown to be better suited: representational content as relational specification over time and space, cf. (Kim, 1996), pp. 200-202. Criteria for this approach are formalised, and it is shown that the formalisation of Damasio's notions indeed fit these criteria.

A brief summary of the main basic assumptions underlying Damasio's approach is expressed in: 'First, I am suggesting that (…) 'having a feeling' is not the same as 'knowing a feeling', that reflection on feeling is yet another step up. (…) The inescapable and remarkable fact about these three phenomena – emotion, feeling, consciousness – is their body relatedness. (…) As the representations of the body grow in complexity and coordination, they come to constitute an integrated representation of the organism, a proto-self. Once that happens, it becomes possible to engender representations of the proto-self as it is affected by interactions with a given environment. It is only then that consciousness begins, only thereafter that an organism that is responding beautifully to its environment begins to discover that *it* is responding beautifully to its environment. But all of these processes – emotion, feeling, and consciousness – depend for their execution on representations of the organism. Their shared essence is the body. (Damasio, 2000), pp. 283-284.

In Section 2 the modelling approach used is briefly introduced. In Sections 3, 4, and 5, for a simple example models are presented for the processes leading to emotion, feeling, and feeling a feeling (or conscious feeling), respectively. Section 6 provides the results of a simulation of these models. In Section 7 it is analysed in how far the representational content of Damasio's notions can be described by two approaches from Philosophy of Mind. Formalisations of some of the dynamic properties of the processes leading to emotion, feeling and feeling a feeling are presented. Next, Section 8 addresses verification. It is shown that the notions for representational content developed in Section 7 indeed hold for the model. The verification is performed both by automated checks and by mathematical proof. Section 9 concludes the paper with a discussion.

2 Modelling Approach

To model the making of emotion, feeling and core consciousness, dynamics play in important role. Dynamics will be described in the next section as evolution of *states* over time. The notion of state as used here is characterised on the basis of an ontology defining a set of state properties that do or do not hold at a certain point in time. The modelling perspective taken is not a symbolic perspective, but essentially addresses the neural processes and their dynamics as neurological processes. This implies that states are just neurological states. To successfully model such complex processes, forms of abstraction are required; for example:

- neural states or activation patterns are modelled as single state properties
- large-dimensional vectors of such (distributed) state properties are composed to one single composite state property, when appropriate; e.g., (p1, p2, …) to p and (S1, S2, …) to S in Section 3.

To describe the dynamics of the processes mentioned above, explicit reference is made to time. Dynamic properties can be formulated that relate a state at one point in time to a state at another point in time. A simple example is the following dynamic property specification for belief creation based on observation:

'at any point in time t1, if the agent observes rain at t1,
then there exists a point in time t2 after t1 such that
at t2 the agent has internal state property s'

Here, for example, s can be viewed as a sensory representation of the rain. To express dynamic properties in a precise manner a language is used in which explicit references can be made to time points and traces: the Temporal Trace Language TTL; cf. (Jonker and Treur, 2002). Here a *trace or trajectory* over an ontology Ont is a time-indexed sequence of states over Ont. The sorted predicate logic temporal trace language TTL is built on atoms referring to, e.g., traces, time and state properties. For example, 'in the internal state of agent A in trace γ at time t property s holds' is formalised by state(γ, t, internal(A)) |= s. Here |= is a predicate symbol in the language, usually used in infix notation, which is comparable to the Holds-predicate in situation calculus. Dynamic properties are expressed by temporal statements built using the usual logical connectives and quantification (for example, over traces, time and state properties).

To be able to perform some (pseudo)-experiments, a simpler temporal language has been used to specify simulation models in a declarative manner. This language (the *leads to* language) enables to model direct temporal dependencies between two state properties in successive states. This executable format is defined as follows. Let α and β be state properties of the form 'conjunction of atoms or negations of atoms', and e, f, g, h, non-negative real numbers. In the *leads to* language the notation $\alpha \twoheadrightarrow_{e, f, g, h} \beta$, means:

If state property α hold for a time interval with duration g,
then after some delay (between e and f) state property β will hold
for a time interval of length h.

For a precise definition of the *leads to* format in terms of the language TTL, see (Jonker, Treur, and Wijngaards, 2003). A specification of dynamic properties in *leads to* format has as advantages that it is executable and that it can often easily be depicted graphically.

In Sections 3, 4, 5 and 6, the *leads to* format has been used to create simulation models of the processes leading to emotion, feeling and core consciousness in terms of neural processes. Given this physical-level model and its dynamic properties, a next step is to assign representational content to (some of) the relevant state properties. For nontrivial cases representational content involves histories of interaction between organism and world (Bickhard, 1993; Jonker and Treur, 2003), and this also shows up in Damasio's theory. To specify and analyse the representational content to a number of state properties of the models and the traces they generate, the more expressive TTL format is used in Section 7. Both formats are used in Section 8.

3 Emotion

First Damasio's notion of *emotion* is addressed. He explains this notion as follows: 'The substrate for the representation of emotions is a collection of neural dispositions in a number of brain regions (…) They exist, rather, as potential patterns of activity arising within neuron ensembles. Once these dispositions are activated, a number of consequences ensue. On the one hand, the pattern of activation represents, within the brain, a particular emotion as 'neural object'. On the other, the pattern generates explicit responses that modify both the state of the body proper and the state of other brain regions. By so doing, the responses create an emotional state, and at that point, an external observer can appreciate the emotional engagement of the organism being observed. (Damasio, 2000), p. 79. According to this description, an *internal emotional state* is a collection of neural dispositions in the brain, which are activated as a reaction on a certain stimulus. Once such an internal emotional state occurs, it entails modification of both the body state and the state of other brain regions. By these events, an *external emotional state* is created, which is accessible for external observation.

Assume that the music you hear is so special that it leads to an emotional state in which you show some body responses on it (e.g., shivers on your back). This process is described by executable local dynamic properties taking into account internal state properties sr(music) for activated sensory representation of hearing the music, and (p1, p2, …) a vector for the activation of preparatory states for the body responses (S1, S2, …); see Figure 1.

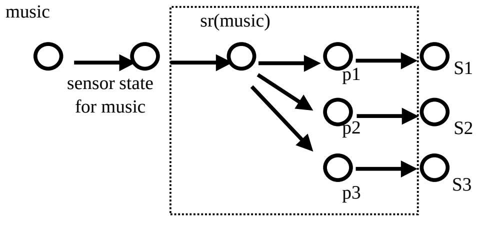

Figure 1: Processes leading to an emotional state

These vectors are the possible internal emotional states. Note that the state properties are abstract in the sense that a state property refers to a specific neural activation pattern. In the model the conjunction p1 & p2 & .. of these preparatory state properties is denoted by p; this p can be considered a composite state property. Moreover, the conjunction of the vector of all body state properties responding to the music S1, S2, … (i.e., the respective body state properties for which p1, p2, ... are preparing) is denoted by (composite) state property S.

The model abstracted in this manner is depicted in Figure 2, upper part. In formal textual format these local properties are as follows:

LP0 music •↠ sensor_state(music)
LP1 sensor_state(music) •↠ sr(music)
LP2 sr(music) •↠ p
LP3 p •↠ S

In the remainder of this paper this abstract type of modelling will be used. Notice, however, that each of the abstract state properties used are realised in the organism in a distributed manner as a large-dimensional vector of more local (neural) state properties. Also the sensory representation sr(music) may be considered such a composite state property with different aspects of the music represented in different forms at different places. Notice, moreover, that the names of the state properties have been chosen to support readability for humans. But in principle these names should be considered as neutral indications of neural states, such as n1, n2, and so on.

4 Feeling

Next, Damasio's notion of *feeling* is considered. He expresses the emergence of feeling as follows: As for the internal state of the organism in which the emotion is taking place, it has available both the emotion as neural object (the activation pattern at the induction sites) and the sensing of the consequences of the activation, a feeling, provided the resulting collection of neural patterns becomes images in mind. (…) The changes related to body state are achieved by one of two mechanisms. One involves what I call the 'body loop'. (…) .. the body landscape is changed and is subsequently represented in somatosensory structures of the central nervous system, from the brain stem on up. The change in the representation of the body landscape can partly be achieved by another mechanism, which I call the 'as if body loop'. In this alternate mechanism, the representation of body-related changes is created directly in sensory body maps, under the control of other neural sites, for instance, the prefrontal cortices. It is 'as if' the body had really been changed but it was not. (…) Assuming that all the proper structures are in place, the processes reviewed above allow an organism to undergo an emotion, exhibit it, and image it, that is, feel the emotion. (Damasio, 2000), pp. 79-80. Thus, a feeling emerges when the collection of neural patterns contributing to the emotion lead to mental images. In other words, the organism senses the consequences of the internal emotional state. Damasio distinguishes two mechanisms by which a feeling can be achieved:

1) Via the *body loop*, the internal emotional state leads to a changed state of the body, which subsequently, after sensing, is represented in

somatosensory structures of the central nervous system.

2) Via the *as if body loop*, the state of the body is not changed. Instead, on the basis of the internal emotional state, a changed representation of the body is created directly in sensory body maps. Consequently, the organism experiences the same feeling as via the body loop: it is ‘as if’ the body had really been changed but it was not.

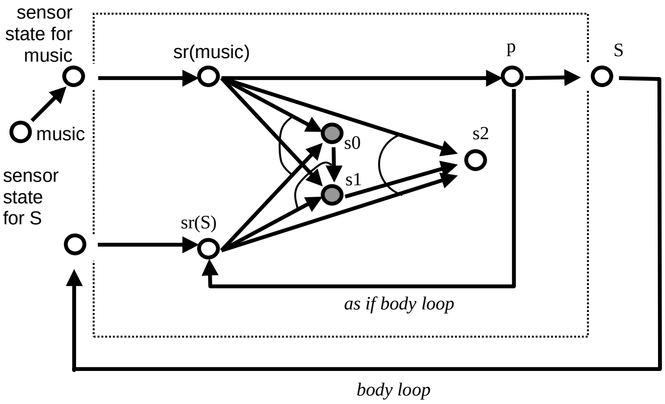

Figure 2: Overview of the simulation model

The model described in Section 3 can be extended to include a number of internal state properties for sensory representations of body state properties that are changed due to responses on the music; together these sensory representations constitute the feeling induced by the music. In Figure 2 the conjunction of these sensory representations is depicted: sr(S) (a sensory representation of the changed body state; this may be materialised in a distributed manner as a kind of vector). This describes the ‘body loop’ for the responses on the music; here S and sensor_state(S) are effects and sensors in the body, respectively. In formal format, two additional local dynamic properties are needed (see also Figure 2):

LP4 S •↠ sensor_state(S)
LP5 sensor_state(S) •↠ sr(S)

Notice that an internal state property sr(shivering) for shivering only, does not directly relate to the music. It is caused by the external stimulus shivering, which in this particular case is originally caused by the music. This body state property shivering could be present for a lot of other reasons as well, e.g., a cold shower. However, taking into account that not only shivering but a larger number of sensory state properties constitute the overall composite state property sr(S), the feeling will be more unique for the music. For the case of an ‘as if body loop’ dynamic properties LP3, LP4 and LP5 can be replaced by the following local dynamic property directly connecting p and sr(S).

LP6 p •↠ sr(S)

Also a combination of models can be made, in which some effects of hearing the music is caused by a body loop and some are caused by an ‘as if body loop’.

5 Feeling a Feeling

Finally, Damasio’s notion of *knowing* or *being conscious of* or *feeling* a *feeling* is addressed. This notion is based on the organism detecting that its representation of its own (body) state (the *proto-self*) has been changed by the occurrence of a certain object (the music in our example). According to Damasio, the proto-self is “a coherent collection of neural patterns which map, moment by moment, the state of the physical structure of the organism”. (Damasio, 2000), p. 177. He expresses the way in which the proto-self contributes to a conscious feeling in the following hypothesis: Core consciousness occurs when the brain’s representation devices generate an imaged, nonverbal account of how the organism’s own state is affected by the organism’s processing of an object, and when this process enhances the image of the causative object, thus placing it in a spatial and temporal context. (p. 169)… with the license of metaphor, one might say that the swift, second-order nonverbal account narrates a story: *that of the organism caught in the act of representing its own changed state as it goes about representing something else*. But the astonishing fact is that the knowable entity of the catcher has just been created in the narrative of the catching process. (…) You know it is *you* seeing because the story depicts a character – you – doing the seeing. (pp. 170-172) … beyond the many neural structures in which the causative object and the proto-self changes are separately represented, there is at least one other structure which *re-represents* both proto-self and object in their temporal relationship and thus represent what is actually happening to the organism: *proto-self at the inaugural instant; object coming into sensory representation; changing of inaugural proto-self into proto-self modified by object*. (p. 177; italics in the original). In summary, the conscious feeling occurs when the organism detects the transitions between the following moments:

1. The proto-self exists at the inaugural instant.
2. An object comes into sensory representation.
3. The proto-self has become modified by the object.

For our case we restrict ourselves to placing the relevant events in a temporal context. In a detailed account, in the trace considered subsequently the following events take place: no sensory representations for music and S occur, the music is sensed, the sensory representation sr(music) is generated, the preparation representation p for S is generated, S occurs, S is sensed, the sensory representation sr(S) is generated. According to Damasio (2000), pp. 177-183, two transitions are relevant (see Damasio's Figure 6.1), and have to be taken into account in a model:

- from the sensory representation of the initial no S body state and not hearing the music to hearing music and a sensory representation of the music, and no S sensory representation
- from a sensory representation of the music and no sensory representation of S to a sensory representation of S and a sensory representation of the music

These two transitions are to be detected and represented by the organism. To model this process three internal state properties are introduced: s0 for encoding the initial situation, and s1 and s2 subsequently for encoding the situations after the two relevant changes. By making these state properties persistent they play the role of indicating that in the past a certain situation has occurred. Local dynamic properties that relate these additional internal state properties to the others can be expressed as follows (see also Figure 2):

LP7 not sr(music) & not sr(S) $\bullet\!\!\twoheadrightarrow$ s0
LP8 sr(music) & not sr(S) & s0 $\bullet\!\!\twoheadrightarrow$ s1
LP9 sr(music) & sr(S) & s1 $\bullet\!\!\twoheadrightarrow$ s2

State properties s0 and s1 are persistent.

6 Simulation

A special software environment has been created to enable the simulation of executable models (Bosse et al., 2004). Based on an input consisting of dynamic properties in *leads to* format (and their timing parameters e, f, g, h, see Section 2), this software environment generates simulation traces. The algorithm used for the simulation is rather straightforward: at each time point, a bound part of the past of the trace (the maximum of all g values of all rules) determines the values of a bound range of the future trace (the maximum of f + h over all LEADSTO rules). The software was written in SWI-Prolog/XPCE, and consists of approximately 20000 lines of code. For more implementation details, see (Bosse et al., 2004).

Using this software environment, the model described in the previous sections has been used to generate a number of simulation traces. An example of such a simulation trace can be seen in Figure 3. Here, 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 executable local properties (i.e., LP0 to LP9), except LP6. In all properties, the values (0,0,1,1) have been chosen for the timing parameters e, f, g, and h. Figure 3 shows how the presence of the music first leads to an emotion (p or S), then to a feeling (sr(S)), and finally to the birth of core consciousness (s2), involving a body loop.

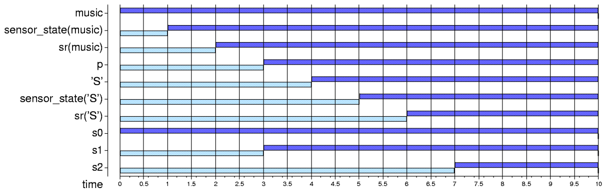

Figure 3: Simulation trace involving a body loop

A similar trace is given in Figure 4, for the case of the as-if body loop. This trace is based on all executable local properties (i.e., LP0 to LP9), except LP3, LP4, and LP5. Again, in all properties, the values (0,0,1,1) have been chosen for the timing parameters e, f, g, and h. As can be seen in Figure 4, in this case the feeling (sr(S)) immediately follows the preparatory state p, without an actual change in body state (S).

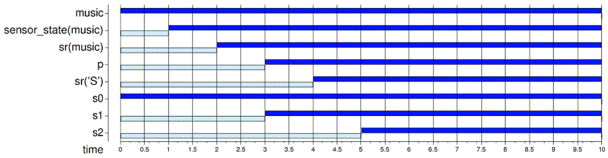

Figure 4: Simulation trace involving an as-if body loop

7 Representational Content

In Damasio's description various types of representation are used, for example, sensory representations and second-order representations. In the literature on Philosophy of Mind a number of approaches to representational content are discussed. In this section three of these approaches are briefly introduced and it is discussed in how far the types of representation used by Damasio indeed can be considered as such according to these approaches.

In (Kim, 1996), pp. 191-192 the *causal/correlational approach* to representational content is explained as follows. Suppose that, some causal chain is connecting an internal state property s and external state property 'horse nearby'. Due to this causal chain, under normal conditions internal state property s of an organism covaries regularly with the presence of a horse: this state property s occurs precisely when a horse is present nearby. Then the occurrence of s has the presence of the horse as its representational content. Especially for perceptual state properties this may work well.

In (Kim, 1996), pp. 200-202 the concept of *relational specification* of a state property is put forward as an approach to representational content. It is based on a specification of how an internal state property can be related to properties of states distant in space and time. This approach is more liberal than the causal/correlational approach, since it is not restricted to one external state, but allows reference to a whole sequence of states in history.

Finally, the *temporal-interactivist approach* (Bickhard, 1993) relates the occurrence of internal state properties to sets of past and future interaction traces. Thus, like the relational specification approach, this approach allows reference to a whole sequence of states in history (or future). However, whilst in the relational specification approach these states can have any desired type (e.g., internal, external, or interaction states), in the temporal-interactivist approach they are restricted to interaction states (i.e. observations and actions).

In the following sections it is explored whether these approaches can be used to specify the representational content of the relevant mental states that occur in our model (i.e., the states that represent emotion, feeling, and feeling a feeling). The focus is on the causal/correlational approach and the relational specification approach. The temporal-interactivist approach is not discussed. However, the formulae expressing the representational content according to the relational specification approach can be easily translated to the temporal-interactivist approach by replacing the external states that occur in the formulae by interaction states (e.g., replacing music by sensor_state(music)).

7.1 Content of Emotion

Consider the causal chain music - sensor_state(music) - sr(music) - p - S (see Figure 1). Thus, looking backward in time, the external emotional state property S can be considered to (externally) represent the emotional content of the music. On the other hand, the internal emotional state property involved is p. Given the causal chain above the (backward) representational content for both p and S is the presence of this very special music, which could be considered acceptable. However, following the same causal chain, also the state property sr(music) has the same representational content. What is different between p and sr(music)? Why are the emotional responses to the same music different between different individuals? This would not be explainable if in all cases the same representational content is assigned. It might be assumed that state properties such as sr(music) may show changes between different individuals. However, the differences are probably much larger between the ways in which for two different individuals sr(music) is connected to a composite state property p. This subjective aspect is not taken into account in the causal/correlational approach. The content of such an emotional response apparently is more personal than a reference to an objective external factor, so to define this representational content both the external music and the internal personal make up has to be taken into account.

For the relational specification approach the representational content of p can be specified in a manner similar to the causal/correlational approach by ‘p occurs if the very special music just occurred’, and conversely. However, other, more suitable possibilities are available as well, such as, ‘p occurs if the very special music just occurred, and by this organism such music was perceived as sr(music) and for this organism sr(music) leads to p’, and conversely. This relational specification involves both the external music and the internal make up of the organism, and hence provides a subjective element in the representational content, in addition to the external reference. This provides an explanation of differences in emotional content of music between individuals.

7.2 Content of Feeling

The representational content of sr(S) according to the causal/correlational approach can consider the causal chain music - sensor_state(music) - sr(music) - p - S - sensor_state(S) - sr(S). Using this chain, sr(S) can be related to both the presence of S, and further back to the presence of the very special music. This steps outside the context of having a reference to one state, which limits the causal/correlational approach. A more suitable approach is the relational specification approach, which allows such temporal relationships to different states in the past; there is the following temporal relation between the occurrence of sr(S), the presence of the S, and the presence of music: ‘sr(S) occurs if S just occurred, preceded by the presence of the music’, and conversely.

7.3 Content of Feeling a Feeling

The representational content of s0 according to the causal/correlational approach can be taken as the absence of both S and music in the past, via the causal chain: no S and no music - sensor state no S and sensor state no music - no sr(music) and no sr(S) - s0. This can be expressed relationally by referring to one state in the past: ‘if no S and no music occur, then later s0 will occur,’ and conversely. Formally:

$$\begin{array}{l} \forall t1\ \ [\ \mathrm{state}(\gamma,\ t1,\ EW)\ |\!==\ \neg\ S \wedge \neg\ \mathrm{music}\ \ \Rightarrow \\ \quad \exists t2 \geq t1\ \ \mathrm{state}(\gamma,\ t2,\ \mathrm{internal})\ |\!==\ s0\ \ ] \\ \forall t2\ \ [\mathrm{state}(\gamma,\ t2,\ \mathrm{internal})\ |\!==\ s0\ \ \ \Rightarrow \\ \quad \exists t1 \leq t2\ \ \mathrm{state}(\gamma,\ t1,\ EW)\ |\!==\ \neg\ S \wedge \neg\ \mathrm{music}] \end{array}$$

For s1 and s2 the causal/correlational approach does not work very well because these state properties essentially encode (short) histories of states. For example, the representational content of s1 according to causal/ correlational approach can be tried as follows: presence of the music and no S in the past under the condition that at some point in time before that point in time no music occurred. However, this cannot be expressed adequately according to the causal/ correlational approach since it is not one state in the past to which reference is made, but a history given by some temporal sequence. The problem is that no adequate solution is possible, since the internal state properties should in fact be related to sequences of different inputs over time in the past. This is something the causal/correlational approach cannot handle, as reference has to be made to another state at one time point, and it is not possible to refer to histories, i.e., sequences of states over time, in the past. A better option is provided by representational content of s1 as relational specification: 'if no S and no music occur, and later music occurs and still no S occurs, then still later s1 will occur,' and conversely. Formally:

$$
\begin{array}{l}
\forall t1, t2\ [\ t1 \leq t2\ \ \&\ \ \mathrm{state}(\gamma, t1, EW) \mid== \neg\, S \wedge \neg\, music\ \ \& \\
\quad \mathrm{state}(\gamma, t2, EW) \mid== \neg\, S \wedge music\ \ \Rightarrow \\
\qquad \exists t3 \geq t2\ \ \mathrm{state}(\gamma, t3, internal) \mid== s1\ \ ] \\
\forall t3\ \ [\ \mathrm{state}(\gamma, t3, internal) \mid== s1\ \Rightarrow\ \exists t1, t2\quad t1 \leq t2 \leq t3\ \ \& \\
\quad \mathrm{state}(\gamma, t1, EW) \mid== \neg\, S \wedge \neg\, music\ \ \& \\
\quad \mathrm{state}(\gamma, t2, EW) \mid== \neg\, S \wedge music\ \ ]
\end{array}
$$

Similarly, the representational content of s2 as relational specification can be specified as follows: 'if no S and no music occur, and later music occurs and still no S occurs, and later music occurs and S occurs, then still later s2 will occur,' and conversely. Formally:

$$
\begin{array}{l}
\forall t1, t2, t3\ \ [\ \ t1 \leq t2 \leq t3\ \ \&\ \ \mathrm{state}(\gamma, t1, EW) \mid== \neg\, S \wedge \neg\, music\ \ \& \\
\quad \mathrm{state}(\gamma, t2, EW) \mid== \neg\, S \wedge music\ \ \& \\
\quad \mathrm{state}(\gamma, t3, EW) \mid== S \wedge music\ \ \Rightarrow \\
\qquad \exists t4 \geq t3\ \ \mathrm{state}(\gamma, t4, internal) \mid== s2\ \ ] \\
\forall t4\ \ [\ \ \mathrm{state}(\gamma, t4, internal) \mid== s2\ \ \Rightarrow \\
\quad \exists t1, t2, t3\quad t1 \leq t2 \leq t3 \leq t4\ \ \& \\
\quad \mathrm{state}(\gamma, t1, EW) \mid== \neg\, S \wedge \neg\, music\ \ \& \\
\quad \mathrm{state}(\gamma, t2, EW) \mid== \neg\, S \wedge music\ \ \& \\
\quad \mathrm{state}(\gamma, t3, EW) \mid== S \wedge music]
\end{array}
$$

This comes close to the transitions mentioned in Section 5: *the proto-self exists at the inaugural instant - an object comes into sensory representation - the proto-self has become modified by the object.*

The above relational specification is a first-order representation in that it refers to external states of world and body, whereas Damasio's second-order representation refers to internal states (other, first-order, representations) of the proto-self. The relational specification given above only works for body loops, not for 'as if body loops'. A relational specification that comes more close to Damasio's formulation, and also works for 'as if body loops' is the following (**RSP**):

$$
\begin{array}{l}
\forall t1, t2, t3\ \ [\ \ t1 \leq t2 \leq t3\ \ \& \\
\quad \mathrm{state}(\gamma, t1, internal) \mid== \neg\, sr(S) \wedge \neg\, sr(music)\ \ \& \\
\quad \mathrm{state}(\gamma, t2, internal) \mid== \neg\, sr(S) \wedge sr(music)\ \ \& \\
\quad \mathrm{state}(\gamma, t3, internal) \mid== sr(S) \wedge sr(music)\ \ \Rightarrow \\
\qquad \exists t4 \geq t3\ \ \mathrm{state}(\gamma, t4, internal) \mid== s2\ \ ] \\
\forall t4\ \ [\ \ \mathrm{state}(\gamma, t4, internal) \mid== s2\ \ \Rightarrow \\
\quad \exists t1, t2, t3\quad t1 \leq t2 \leq t3 \leq t4\ \ \& \\
\quad \mathrm{state}(\gamma, t1, internal) \mid== \neg\, sr(S) \wedge \neg\, sr(music)\ \ \& \\
\quad \mathrm{state}(\gamma, t2, internal) \mid== \neg\, sr(S) \wedge sr(music)\ \ \& \\
\quad \mathrm{state}(\gamma, t3, internal) \mid== sr(S) \wedge sr(music)\ \ ]
\end{array}
$$

This is a relational specification in terms of other representations (sr(music), sr(S)), and therefore a second-order representation. It has no direct reference to external states anymore. However, indirectly, via the first-order representations sr(music) and sr(S) it has references to external states.

8 Verification

In Sections 3-6, local, executable dynamic properties were addressed, and simulation based on these properties was discussed. In Section 7, dynamic properties to describe representational content of internal states are introduced. These dynamic properties are of a *global* nature. Another example of a more global property is the following:

OP1 music $\bullet\!\!\twoheadrightarrow$ s2

Informally, this property states that the presence of music eventually leads to the birth of core consciousness (s2). This can be considered as a global property because it describes dynamic of the overall process, whereas the properties presented in Sections 3-6 described basic steps of the process. For both types of global properties (i.e., dynamic property OP1 and the properties specifying representational content), an important issue is *verification*. In other words, are these global properties satisfied by the simulation model described in Sections 3-6? Therefore, the global properties have been formalised, and verification has been applied in two ways: by *automated checks* and by establishing *logical relationships*.

8.1 Automated Checks

In addition to the simulation software described in Section 6, a software environment has been developed that enables to check dynamic properties specified in TTL against simulation traces. This software environment takes a dynamic property and one or more (empirical or simulated) traces as input, and checks whether the dynamic property holds for the traces. Using this environment, the global properties mentioned above have been automatically checked against traces like depicted in Figure 3 and 4. The duration of these checks varied between 0.5 and 1.5 seconds, depending on the complexity of the formula. All these checks turned out to be successful, which validates (for the given traces at least) our choice for the representational content of the internal state properties. However, note that these checks are only an empirical validation, they are no exhaustive proof as, e.g., model checking is.

8.2 Logical Relationships

A second way of verification is to establish logical relationships between global properties and local properties. This has been performed in a number of cases. For example, to relate OP1 to local properties, intermediate properties were identified in the form of the following milestone properties that split up the process in three phases:

MP1(MtoE) $\text{music} \bullet\!\!\twoheadrightarrow \text{sr(music)}$ &
$\text{sr(music)} \bullet\!\!\twoheadrightarrow \text{S}$
MP2(EtoF) $\text{S} \bullet\!\!\twoheadrightarrow \text{sr(S)}$
MP3(FtoFF) **RSP** (see Section 7)

For the milestone properties the following relationships hold (for simplicity neglecting ‘as if body loops’):

MP1(MtoE) & MP2(EtoF) & MP3(FtoFF) $\Rightarrow$ OP1
LP0 & LP1 & LP2 & LP3 $\Rightarrow$ MP1(MtoE)
LP4 & LP5 $\Rightarrow$ MP2(EtoF)
LP7 & LP8 & LP9 $\Rightarrow$ MP3(FtoFF)

Figure 5 provides the same relationships in the form of a logical AND-tree.

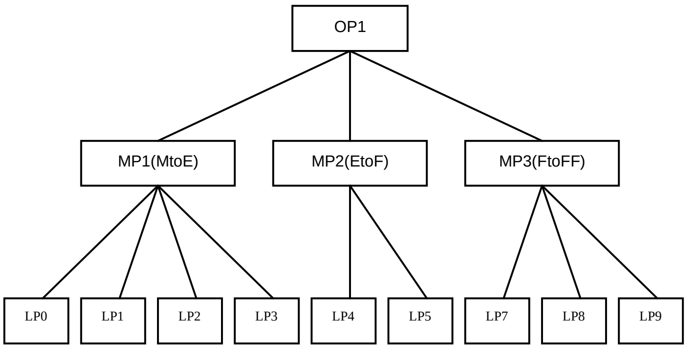

Figure 5: Logical relationships between the dynamic properties

Such logical relationships between properties can be very useful in the analysis of traces. For example, if a given trace that is unsuccessful does not satisfy milestone property MP2, then by a refutation process it can be concluded that the cause can be found in either LP4 or LP5. In other words, either the sensor mechanism fails (LP4), or the sensory representation mechanism fails (LP5).

9 Discussion

The chosen modelling approach describes temporal dependencies in processes at a neurological, not symbolic level. To avoid complexity the model is specified at an abstract level. From the available approaches to representational content from Philosophy of Mind, the causal/correlational approach is not applicable, but Kim’s relational specification approach, that allows more complex temporal dependencies, is applicable. Using this approach, claims on representational content made by Damasio have been formalised and supported by means of verification.

Furthermore, an interesting observation that has been made on the basis of the formalisation was that the model predicted the possibility of ‘false core consciousness’: core consciousness that is attributed to the ‘wrong’ stimulus. To explain this phenomenon, suppose that two stimuli occur, say x1 and x2, where x2 is subliminal and unnoticed. Then, it could be the case that x2 provokes emotional responses, whilst the conscious feeling that arises is attributed to x1 instead of x2. In terms of our model, this can be simulated by first introducing a subliminal stimulus that yields emotion S (e.g., a cold breeze) followed by the stimulus music. In that case, the conscious feeling would incorrectly be attributed to the music. In personal communication with Antonio Damasio, the existence of this predicted false core consciousness was confirmed.

For the philosophical perspective the paper contributes a case study for representational content which is more down-to-earth than the science fiction style thought experiments, such as the planet Twin Earth, that are common in the literature on Philosophy of Mind, e.g., (Kim, 1996). In addition, the type of representation is more sophisticated than the usual ones essentially addressing sensory representations induced by observing (a snapshot of) a horse or a tomato. Interesting further work in this area is to analyse various arguments given in this literature by applying them to this example.

The analysis approach that is applied in this paper to model Damasio’s theory of consciousness, has previously been applied to complex and dynamic cognitive processes other than consciousness, such as the interaction between agent and environment (Bosse, Jonker, and Treur, 2004). In a number of these cases, in addition to simulated traces, also empirical (human) traces have been formally analysed. Using this approach, it is possible to verify global dynamic properties (e.g., specifying the representational content of internal states) in real-world situations.

For recent work in the area of emotion and consciousness, the interested reader is referred to (Prinz and Chalmers, 2004), Chapter 3, which gives an account for emotions as embodied representations of “core relational themes” such as danger and obstruction.

Acknowledgements

The authors are grateful to Antonio Damasio for his valuable comments upon their questions, and to an anonymous referee for some suggestions for improvements of an earlier version of this paper.

References

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

Bosse, T., Jonker, C. M., van der Meij, L., and Treur, J., LEADSTO: a Language and Environment for Analysis of Dynamics by SimulaTiOn. Vrije Universiteit Amsterdam, Department of Artificial Intelligence. Technical Report, 2004.

Bosse, T., Jonker, C.M., and Treur, J., *Representational Content and the Reciprocal Interplay of Agent and Environment* In: Leite, J., Omincini, A., Torroni, P., and Yolum, P. (eds.), Proceedings of the Second International Workshop on Declarative Agent Languages and Technologies, DALT'04, Springer Verlag, 2004, pp. 61-76.

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

Damasio, A., *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. MIT Press, 2000.

Damasio, A., Grabowski, T.J., Bechara, A., Damasio, H., Ponto, L.L.B., Parvizi, J., and Hichwa, R.D., Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, vol. 3, 2000, pp. 1049-1056.

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, C.M. and Treur, J., A Temporal-Interactivist Perspective on the Dynamics of Mental States. *Cognitive Systems Research Journal*, vol. 4, 2003, pp. 137-155.

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.

Kim, J., *Philosophy of Mind*. Westview Press, 1996.

Parvizi, J. and Damasio, A., Consciousness and the brain stem. *Cognition*, vol. 79, 2001, pp. 135-159.

Prinz, J.J. Chalmers, D.J., Gut Reactions: A Perceptual Theory of Emotion (Philosophy of Mind (Hardcover)), Oxford University Press, 2004.