Int. J. Human-Computer Studies (2002) **57**, 165–214
doi:10.1006/ijhc.1023
Available online at http://www.idealibrary.com.on IDEAL®

Modelling multiple mind–matter interaction

CATHOLIJN M. JONKER
Department of Artificial Intelligence, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands. email: jonker@cs.vu.nl, URL: http://www.cs.vu.nl/~jonker

JAN TREUR
Department of Artificial Intelligence, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands.

Department of Philosophy, Utrecht University, Padualaan 8, 3584 CS, Utrecht, The Netherlands. e-mail: treur@cs.vu.nl, URL:http://www.cs.vu.nl/~treur

(Received 15 February 2001 and accepted in revised form 17 July 2002)

Relations between mental and physical aspects of an agent can be of various types. Sensing and acting are among the more commonly modelled types. In agent modelling approaches often this is the only interaction between the physical and mental; other possible types of interactions are abstracted away. If it is also taken into account that the agent's mind has a materialization in the form of a brain, the relations between mind and matter may become more complex. An explanation of a dynamic pattern may involve mental aspects, physical aspects, and interactions between mental and physical aspects. An explanatory perspective sometimes advocated for such more complex phenomena is explanatory pluralism. According to this perspective an explanation can consist of parts of a different signature, for example, a partial physical explanation and a partial mentalistic explanation. Each of these partial explanations is insufficient to explain the whole phenomenon, but together they do explain the whole, if some interaction is assumed. How for such explanations the different types of interaction between mind and matter of an agent and the material world can be modelled in a conceptually and semantically sound manner, and how the overall explanation is composed from the parts, using these interactions, is the main topic of this paper. The generic model presented can be used to model, explain and simulate a variety of phenomena in which multiple mind–matter interactions occur, including, for example, sensing and acting, (planned) birth and death, bacterial behaviour, getting brain damage, psychosomatic diseases and applications of direct brain–computer interfaces.

KEYWORDS: mind; matter; agent; dynamics; pluralist explanation; interaction.

1. Introduction

A widespread traditional view on the use of relationships between cognitive and neurological theories, is that these can be used to reduce and actually replace a cognitive explanation of a phenomenon by a neurological explanation (e.g. Nagel, 1961; Kim,

1071-5819/02/$ - see front matter

1996). From the viewpoint of scientific practice this view is debatable. Two main arguments against this view are as follows.

(1) In scientific practice, where theories are incomplete and still under development, only incomplete inter-theory relationships can be found, and replacement can only take place partially
(2) Even if sufficiently complete inter-theory relationships can be found, explanations from a lower-level theory are often unsatisfactory because they are not transparent or intractable due to the (too) large number of details involved.

As opposed to this traditional view, *explanatory pluralism*, McCauley (1996) claims that in scientific practice explanations from theories at different levels of description, like psychology and neuroscience, both play a role, without an explanation of one level being reduced to or replaced by an explanation of the other level. This position is further extended by Bechtel and Mundale (1999) and McCauley and Bechtel (2001); one of the claims is that assumed heuristic identifications between descriptions at different levels play an important role in directing the research. Explanatory pluralism seems to fit well to scientific practice, for example in Cognitive Neuroscience and Biology, where often explanations from different theories co-exist and interact. As an example, Bechtel and Mundale (1999) show how in animal studies neuro-psychologists exploit the homology between nervous systems of different species to draw conclusions on the functions of human brain areas. Another example, discussed by Schouten and Looren de Jong (2001) and Looren de Jong (2002) is the case of behavioural genetics, where theories from neuro-physiology, personality psychology and molecular genetics interact. Recent developments in molecular genetics have opened at least a possibility of relating the base pair sequence on the genome to personality characteristics. Such a relation is very complex and defies attempts in the philosophy of science to lay down necessary and sufficient conditions for actual reduction of explanations to the lower description level. In these examples the emphasis is on incomplete knowledge as a reason to combine explanations from different description levels.

From another perspective, in Dennett (1987, 1991), as opposed to explanations from a direct physical stance, the *intentional stance* is put forward to provide higher-level explanations. Different description levels with ontologies for emerging patterns in the simulation environment Life are used to explain the advantage of explanations for more complex phenomena using such a higher-level explanation (cf. Dennett, 1987, pp. 37–39, 1991, pp. 37–42). In addition, he uses the description levels in computer systems (actually of a chess computer), embedded (and hence visualized) in the two-dimensional Life environment as a metaphor to explain the advantage of design stance and intentional stance explanations for mental phenomena over physical stance explanations:

> The scale of compression when one adopts the intentional stance toward the two-dimensional chess-playing computer galaxy is stupendous: it is the difference between figuring out in your head what white's most likely (best) move is versus calculating the state of a few trillion pixels through a few hundred thousand generations. But the scale of savings is really no greater in the Life world than in our own. Predicting that someone will duck if you throw a brick at him is easy from the folk-psychological stance; it is and will

> always be intractable if you have to trace the protons from brick to eyeball, the neurotransmitters from optic nerve to motor nerve, and so forth. (Dennett, 1991, p. 42)

Dennett puts the emphasis on tractability. To explain phenomena in our real world, for the simpler processes a physical stance is appropriate, but for more complex processes only higher-level explanations are tractable, whereas lower-level explanations are not. Also Dennett's brick example shows that in one phenomenon both types of explanation can come together. The mechanics of the brick's trajectory can be adequately explained from the physical stance (using the laws of classical mechanics), whereas the ducking needs a higher-level explanation (e.g. from the intentional stance).

In the same line, within Biology the attempt to understand the behaviour of a cell such as *Escherichia coli*, and the dynamics of its intracellular processes in terms of the underlying biochemistry leads to hundreds of differential equations with parameters for which reliable estimations are rarely known. Given that two coupled differential equations already can show complex behaviour, even if all parameters were known, this type of description may be intractable and add no understanding. Even taking into account that in this case the biochemistry is by and large known, this situation may be considered similar to explaining from the physical stance why a person ducks when a brick is thrown, and Dennett's analysis may apply here accordingly. Higher-level descriptions may be more adequate for scientific practice than the lower-level biochemistry descriptions. One approach recognizes that some conglomerates of biochemical processes act as functional units such as "metabolic pathway", "catabolism", "transcriptome" and "regulon". Some of these concepts have been or are being defined formally (Kahn & Westerhoff, 1991; Rohwer, Schuster & Westerhoff, 1996; Schilling, Letscher & Palsson, 2000). Viewed from a more high-level perspective, the cell effectively makes decisions regarding its internal dynamics and externally observable behaviour, given its environmental circumstances, and implements these decisions into appropriate actions. This behaviour, viewed from this high-level perspective is less complex than the hundreds of differential equations. This suggests that considering a cell from the perspective of an agent sensing the environment, integrating that information within its internal state, and then choosing behavioural patterns of action, may provide the basis of an alternative approach to modelling and explanation. Some first steps from such a high-level perspective show promising results (cf. Jonker, Snoep, Treur, Westerhoff & Wijngaards, 2002).

The use of explanations from theories at different description levels in scientific practice as described, for example within explanatory pluralism, raises the question of how actually explanations from different theories interact, given that only incomplete relationships between the theories are assumed. This question is analysed in more detail in this paper for the case of an agent that is partly described from a cognitive perspective (mind) and partly from a physical perspective (matter). Usually, within more traditional Artificial Intelligence or Cognitive Science theories or models to describe an agent from the cognitive perspective isolate the agent's mental functioning (often modelled by *symbolic* means) from the agent's *embedding* in the physical world. No interactions are possible between processes within the world and mental processes, except those interactions defined by sensing and performing actions. This disembodied view on modelling has as a strong *modelling advantage* that no disturbances from the external world have to be taken into account, which makes the model simpler. However, this also entails a strong *disadvantage* in the sense of *a lack of realism*: the concepts within the model cannot be related in a direct manner to the (physical) reality (e.g. Kim, 1996). This problem is sometimes called the *symbol-grounding problem*; see, for example, Sun (2000*a*). Explanation of the agent's behaviour has to take place exclusively within the (symbolic) cognitive description, taking into account solely the interaction by sensing and performing actions. This entails a form of explanatory exclusivism rather than pluralism. In particular, phenomena that depend in a crucial manner on the interaction between the agent's physical and mental description cannot be modelled nor explained. Two types of such interaction may occur

(1) Physical circumstances cause changes in mental functioning, but not via the agent's sensors—e.g. for a human a loss of memory due to a car accident, or for a bacterium a changed intentional behaviour due to DNA that is damaged by radiation.
(2) Mental processes cause changes in the physical world, but not via the agent's effectors—e.g. telekinesis, or, if for physical reasons this is not considered to exist, changes in the physical world on the basis of EEG patterns, as exploited in the area of brain–computer interfacing.

A phenomenon of type (1) or (2) cannot be modelled by a classical approach. For example, if an intentional model of bacterial behaviour is based on the BDI-agent model of Rao and Georgeff (1991), it cannot be described how radiation can affect the presence of desires.

As an alternative, in continuation of the work reported in Jonker and Treur (1997), this paper introduces a *hybrid agent modelling* approach where both the cognitive (symbolic) aspects and the physical aspects and their interaction are covered. This modelling approach is able to model dynamic phenomena as indicated in (1) and (2) above; in particular, the hybrid approach allows one to specify and simulate in an integrated manner.

(1) A model for the agent's *cognitive* processes (as usually modelled by symbolic models e.g. logical models).
(2) A model for the agent's *physical* processes (as usually modelled by mathematical models e.g. connectionist models).
(3) Multiple *interactions* between these two models of four different types:
    (a) Matter–mind interaction via the sensors.
    (b) Mind–matter interactions via the effectors.
    (c) Direct interactions from mind to matter.
    (d) Direct interactions from matter to mind.

Thus the hybrid agent modelling approach subsumes and integrates the two different modelling paradigms. On the basis of such a model, pluralist explanations can be obtained that are partly at the cognitive level and partly at the physical level.

In this paper, first in Section 2 two examples of multiple mind–matter interaction are introduced. These examples will be used as the main illustration throughout the paper. It is shown how explanations of processes within these examples can be based on (combined) traces within the two levels of descriptions (mind and matter) that interact. In Section 3 the concept of combined interacting mind–matter traces is defined in more detail. Section 4 analyses such traces from the perspective of representation. Section 5 briefly introduces the component-based modelling approach DESIRE that is used for the agent model for multiple mind–matter interaction. This generic model is presented in Section 6 in different parts as given below.

(1) The model of the material world and its symbolic representation.
(2) The agent's symbolic representation and its interaction with the material world by sensing and acting.
(3) The agent's (embodied) material representation, and its relation to the agent's symbolic representation.
(4) The embedding of the agent's material representation within the rest of the material world.

Section 7 presents two instantiations of this model for the two case studies introduced in Section 2, and Section 8 presents two simulation traces of these refined models. In Section 9 for a number of other examples of multiple mind–matter interaction it is briefly discussed how the model could be used to describe or simulate them. The paper concludes with a discussion (Section 10).

2. Examples of explanations based on multiple mind–matter interaction

In this section two examples are introduced that illustrate the issue addressed in this paper. Both examples will return later on in the paper as well.

2.1. THE ICECREAM EXAMPLE

Consider an agent walking down a street, see Figure 1 (position p1). The agent observes an icecream sign at the supermarket across the street (the supermarket is at position p3 in Figure 1). As he did not eat or drink for a few hours he has a desire for icecream, so he sets himself the intention of crossing the street. Although the shrub to his left limits his view of the road, he decides to cross the street as he does not see any cars. Unfortunately, there is a car coming down the street. The driver, being a bit in a hurry, comes around the curve with the shrub (position p2 in Figures 1 and 2) at the same moment that the agent arrives at position p2. As can be seen in Figure 2, the car hits the agent. Although the accident is a minor one (the agent has no permanent injuries), the agent is momentarily stunned and suffers from temporary amnesia (his short-term memory is lost). One of the effects is that the agent cannot remember his intention to visit the supermarket. Furthermore, he cannot remember any of the observations he made prior to the crossing of the street. Realizing that he lacks knowledge about his present predicament, the agent decides to observe his surroundings (again).

How can this course of affairs be explained? For the first part of the process, an (iterated) high-level explanation based on the intentional stance is an adequate option:

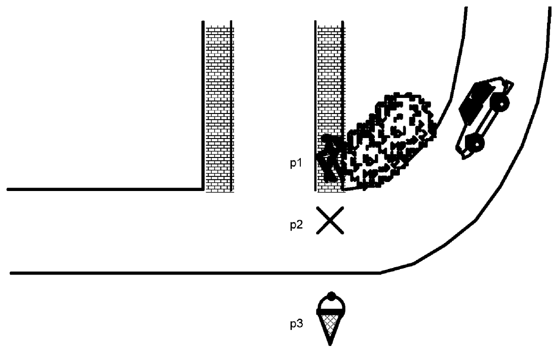

FIGURE 1. Initial situation.

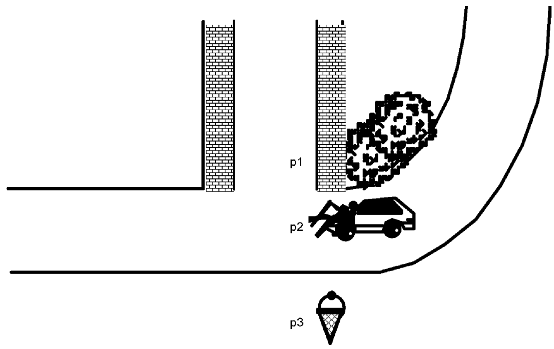

FIGURE 2. Situation at the time of the accident.

Why is the agent crossing the street?
The agent crossed the street because he had the intention to cross the street and he believed that there was an opportunity to do so (i.e. no traffic coming).

Why did the agent have the intention to cross the street, and why did he believe that no traffic was coming?
The agent had the intention to cross the street because he desired to have an icecream and he believed that on the other side of the street he could have icecream.
The agent believed that no traffic was coming because he observed an empty street and did not observe traffic further away.

Why did the agent have the desire for icecream, and why did he believe that at the other side of the street he could have icecream?
The agent had the desire for icecream because he did not eat or drink for a few hours.
The agent believed that on the other side of the street he could have icecream because he observed an icecream sign over there.

This explanation fits well in the intentional stance. An explanation from a physical stance is not at issue for this part of the process; it would be intractable. Note that the observations and the actions connect to the physical world; they are end points in the explanation. But what about the rest of the story:

> *Why does the agent not remember the last few minutes?*
> The agent does not remember the last few minutes because his brain has some damage.
>
> *Why did the agent's brain get damaged?*
> The agent's brain got damaged because he was on the street while the car was passing.
>
> *Why was the agent on the street and why was the car passing at that time point?*
> The agent was on the street because he was crossing the street.
> The car was passing there at that point in time because on its way from A to B, it was passing this point at this time.

This explanation is an explanation from a physical stance. Two material bodies were following physical trajectories that brought both at the same place at the same time with the car at a high speed. This caused physical damage in the agent's body, including his brain. The brain damage caused bad functioning of memory for a few minutes back in time. Note that the first part of this explanation actually is an interaction between the higher- and the lower-level descriptions: the higher-level notion "not remembering" is connected to the lower-level notion "having some type of brain damage". A similar connection is made between the higher-level notion "crossing the street" (as an action initiated by the agent) and the lower-level notion "being on the street" (the action's effect within the material world). This is where the two partial explanations interact. This interaction is the glue used to compose the two to a composite or pluralist explanation of the whole phenomenon.

2.2. THE E. COLI EXAMPLE

For the well-known bacterium *E. coli* to interact with its external environment, several mechanisms are known. One of these is an observation mechanism to find out which food substances are present in the environment. In the simplified case we consider, we will address only glucose or lactose. Such mechanisms also enable *E. coli* to actively import a food substance (of its choice). This is a way of performing actions. *E. coli* makes the choice between the two types of food in the following manner.

> If glucose is present in the environment, then it will import glucose and it will not import lactose.
> If lactose is present in the environment, but not glucose, then it will import lactose.

This can be summarized and interpreted as: *E. coli* can use both substances, but it prefers glucose; it never imports both. In particular, this means that in an environment where first only lactose is present, and next glucose is added such that both are present, it first imports lactose, but later on it stops the lactose import and starts glucose import.

An example of a physical stance explanation based on the bacterium's chemistry is as follows:

> *Why does this E. coli import lactose?*
> This E. coli imports lactose because the concentration of CRPcAMP was above 0.01 mmol/l, and the concentration of the lactose import enzyme was above 0.1 mmol/l.
>
> *Why was the concentration of CRPcAMP above 0.01 mmol/l, and why was the concentration of the lactose import enzyme above 0.1 mmol/l?*
> The concentration of CRPcAMP was above 0.01 mmol/l because the concentration of glucose in the environment was at most 0.1 mmol/l.
> The concentration of the lactose import enzyme was above 0.1 mmol/l because the concentration of lactose import mRNA was above 0.1 mmol/l, and the concentration of lactose internally was above 0.1 mmol/l.
>
> *Why was the concentration of lactose import mRNA above 0.1 mmol/l, and why the concentration of lactose internally was above 0.1 mmol/l?*
> The concentration of lactose import mRNA was above 0.1 mmol/l because the concentration of lactose internally was above 0.1 mmol/l, because the concentration of CRPcAMP was above 0.01 and because of the presence of its DNA.
> The concentration of lactose internally was above 0.1 mmol/l because the concentration of lactose externally was above 0.1 mmol/l.

An intentional stance explanation for the same pattern is as follows:

> *Why does this E. coli import lactose?*
> This E. coli imports lactose because it believed that lactose was present and glucose is not present in the environment, and it had the intention to import lactose.
>
> *Why the intention to import lactose and why these beliefs?*
> It had the intention to perform lactose import because it had the desire to perform lactose import and it believed that lactose is present and no glucose is present in the environment.
> It had the belief that lactose was and glucose is not present externally because it observed that.
>
> *Why the desire to import lactose and why the belief that lactose was present?*
> It had the desire to import lactose because it is an innate desire present from its birth.
> The belief that lactose was present was because the bacterium observed that.

Note that the observations connect to the physical world. In contrast to the icecream case both explanations may be feasible, and can be related to each other (Jonker, Snoep, Treur, Westerhoff & Wijngaards, 2001, 2002). However, an advantage of the higher-level explanation is that it is more intuitive and more generic than the lower-level chemical explanation.

Suppose the higher-level explanation is taken as the preferred one. In another scenario, an accident happens to this *E. coli.* First both lactose and glucose are present in the environment. The bacterium imports glucose. Next, by strong radiation the part of the DNA for glucose import (i.e. the glucose import gene) is damaged, leaving the rest unimpaired. The *E. coli* stops glucose import; it now does not import any nutrient. Next glucose is taken away from its environment. It starts to import lactose. How can all this be explained?

> *Why did this E. coli not import lactose or glucose in the environment with both available?*
> It did not import lactose or glucose because it did not have the intention to import any of these.
>
> *Why not the intention to import neither lactose nor glucose?*
> It had no intention to perform lactose import because it believed that glucose is present in the environment, although having the desire to perform lactose import.
> It had no intention to perform glucose import because it did not have the desire to perform glucose import, although it believed that glucose is present in the environment.
>
> *Why the belief that glucose was present and why not the desire to import glucose?*
> The belief that glucose was present was because the bacterium observed that.
> It had no desire to import glucose because this desire was innate but its gene was damaged by strong electro-magnetic fluctuations.
>
> *Why were strong electro-magnetic fluctuations present in the E. coli?*
> Strong electro-magnetic fluctuations were present because a source of strong radiation was present nearby.

The last part of this explanation is from a physical stance, whereas the first part is from an intentional stance. Halfway a connection is made between the higher-level notion "absence of the desire" and the lower-level notion "damaged gene"; this is where both partial explanations interact.

2.3. INTERACTING TRACES AND COMPOSITE EXPLANATIONS

The pattern that the two pluralist explanations discussed in Sections 2.1 and 2.2 have in common can be schematically depicted as in Figure 3. Two description levels are involved.

(1) First functioning can be explained from the higher-level description.
(2) This leads to changes in the lower-level description (interaction from higher- to lower-level description).
(3) Some further processes take place that are explained by the lower-level description.
(4) At some point in time this lower-level process description interacts with the higher-level description.
(5) After this point in time the functioning can be explained from the higher-level description plus the effect of the interaction with the lower-level description.

The overall explanation is not limited to one of the two description levels, but exploits both levels, and interaction between the levels. This interaction forms the

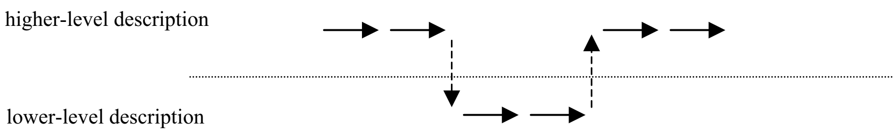

FIGURE 3. An interacting trace as a basis for a composite explanation.

composition or connection principle by which the two parts are glued together to a composite explanation of the whole (Figure 3).

3. Tracing the dynamics of mind–matter interaction

In this section a semantic formalization of the interacting mind–matter traces that play a role in pluralist explanations is presented. To define semantics of the dynamics of the interaction process, temporal models are used; for other uses of this approach, resp. to meta-level architectures, nonmonotonic reasoning processes and component-based reasoning models, see Treur (1994), Engelfriet and Treur (1995) and Brazier, Treur, Wijngaards and Willems (1999). Within this approach the semantics of a process is formalized by a set of temporal models (i.e. sequences of states) that formalize the alternative patterns of the dynamics. The following types of transitions between two subsequent states are allowed in these temporal models:

Single representation transitions
Material state transition
A change in the state of the material description level.

Mental state transition
A change in the state of the mental description level.

Interaction transitions
Downward transduction transition
A change of the material state under the influence of the mental agent state.

Upward transduction transition
A change of the mental agent state under the influence of the material state.

The more precise definitions are given below.

Definition 1 (state): An *information type* Σ is a set of (e.g. propositional or predicate logic) symbols, used as a vocabulary to define a set of *ground atoms* $\mathrm{At}(\Sigma)$. A *state* for an information type Σ is a mapping $\mathsf{M} : \mathsf{At}(\Sigma) \rightarrow \{0, 1\}$ from the set of ground atoms $\mathrm{At}(\Sigma)$ to the set of truth-values $\{0, 1\}$, i.e. a *model*. The set of all states of information type Σ is denoted by $\mathsf{S}(\Sigma)$.

An example of a structure that defines an information type is a tuple of (sub-) sorts, constants, functions and predicates of an order-sorted predicate logic.

Definition 2 (transition): A *transition between states* is a pair of models, i.e. an element $\langle \mathsf{S}, \mathsf{S}' \rangle$ (also denoted by $\mathsf{S} \rightarrow \mathsf{S}'$) of $\mathsf{S}(\Sigma) \times \mathsf{S}(\Sigma)$. A *transition relation* is a relation on $\mathsf{S}(\Sigma) \times \mathsf{S}(\Sigma)$.

Behaviour is the result of transitions from one state to another. If a transition relation is functional, then it specifies deterministic behaviour. By applying transitions in succession, sequences of states are constructed. These sequences, also called traces (and interpreted as temporal models), formally describe behaviour.

Definition 3 (trace): A *trace* or *temporal model* of information type Σ is a sequence of states $(M^t)_{t \in N}$ in $S(\Sigma)$. The set of all temporal models is denoted by $S(\Sigma)^N$, or $Traces(\Sigma)$.

A set of temporal models is a declarative description of the semantics of the behaviour of a process; each temporal model can be seen as one of the alternatives for the behaviour. Next these notions are applied to the two types of description distinguished in a multiple mind–matter interaction process.

Definition 4 (composite state): The *set of composite states* of the whole process is defined by $S = S(\Sigma^{mat}) \times S(\Sigma^{men})$. Here, Σ^{mat} is an information type specifying a vocabulary to represent the material state, Σ^{men} the same for the mental state.

Transitions and traces adhere to the structure of the states: a typed transition describes a composite state that changes in time. Following the component-based structure, only some types of transitions are allowed. As put forward informally above, for each of the description levels a transition limited to this level (leaving untouched the other level) is possible: a *material state change* or a *mental state change* step. Moreover, transitions involving interaction between description levels are *upward transduction* and *downward transduction*. The following definition postulates that (only) these types of transitions are possible.

Definition 5 (typed transition): (a) The following types of transitions are defined:

material state transition $S(\Sigma^{mat}) \rightarrow S(\Sigma^{mat})$
mental state transition $S(\Sigma^{men}) \rightarrow S(\Sigma^{men})$
upward transduction transition (observation) $S(\Sigma^{mat}) \times S(\Sigma^{men}) \rightarrow S(\Sigma^{men})$
downward transduction transition (action) $S(\Sigma^{mat}) \times S(\Sigma^{men}) \rightarrow S(\Sigma^{mat})$
upward transduction transition (other) $S(\Sigma^{mat}) \times S(\Sigma^{men}) \rightarrow S(\Sigma^{men})$
downward transduction transition (other) $S(\Sigma^{mat}) \times S(\Sigma^{men}) \rightarrow S(\Sigma^{mat})$

(b) A *typed transition* is a transition $S \rightarrow S^1$ which is based on a transition of one of the types defined in (a).

The model as described in this paper only makes these types of transitions. Notice that the upward and downward transduction transitions are classified into two further types, depending on whether they are associated to standard agent–environment interactions such as observations and action execution, or not.

Definition 6 (coherent trace): (a) A *typed trace* is a sequence of states $(M^t)_{t \in N}$ in S. The set of all typed traces is denoted by S^N, or $Traces$.

(b) An element $(M^t)_{t \in N} \in$ Traces is called *coherent* if for all time points t the step from M^t to M^{t+1} is defined in accordance with a typed transition. The set of coherent typed traces forms a subset $CTraces$ of $Traces$.

It is possible and in applications often necessary to define more constraints on the transitions. For example, physical laws for the material description level, or mental laws for the mental description level; e.g. if a mental state (reasoning) transition adds information to the mental state, then this information is in the deductive closure of the knowledge (consisting of the knowledge base and the information available to the current mental state).

Note that in the definitions above it is allowed that transduction takes time. However, if desired, as in a slight modification, in an alternative constraint it can be expressed in Definition 6 that a transduction transition works in an instantaneous manner, i.e. it is a relation between different states at one time point. This means that a previous state transition, for example at the material level, although it is defined as (only) changing the material state, as an implication it changes the mental state at the same time. Since conceptually this is less transparent, our choice has been to distinguish this process as a two-steps process in which the second step takes virtually no time. In the next sections examples of traces will be discussed.

3.2. TRACES FOR THE TWO CASES

In this section it is shown how the course of events in the first example introduced in Section 2 is described by interacting traces in Table 1. The trace is started at the moment that the agent is in position p1 and has observed that a supermarket with an icecream sign at position p3, and that a path from p1 to p3 is available with p2 as next position. Moreover, the agent has observed that no car was present. As a result the observation information is available within the agent (as current beliefs). The trace is started at time point t1. The situation at time t1 is represented in Figure 1. In the state description only the state properties that just have been changed are listed.

For the *E. coli* case the trace as depicted in Table 2 can be made. Note that for the icecream case for the generated beliefs, intentions and action initiations, physical representations are included within the trace. The reason for this is that later on the brain damage occurs that changes the material state. In the *E. coli* case the physical representations of beliefs, intentions and actions play no role further on (only the physical representation of the desire plays a role); therefore, they have been left out of the trace. This shows the use of pluralist explanations. For those parts of the process where no physical description is needed, a mental description suffices, whereas for the cases where a physical description is crucial to capture the whole process, such a physical description is included.

4. Representation and simulation in mind–matter interaction

Modelling situations in which interactions between mind and matter occur is not an exception: sensing and acting are among the more commonly modelled types of interaction between mental and physical aspects. However, as mental processes themselves are assumed to be embodied by physical processes, also relations between mental and physical aspects of a different type can occur. In modelling approaches these other possible types interactions are usually abstracted away: the model for the mental

TABLE 1
Example mind–matter trace for the icecream case

State no.	Mental state	Process or interaction	Material state
0	Desiring icecream		Being aside the street Having image of icecream sign in eye
		Upward observation transduction: interaction from material to mental Observation of supermarket and icecream	
1	Observed there is icecream across the street		
		Downward transduction: *interaction from mental to material* Physical representation in the brain	
2			Physical representation of observed facts in the brain
		Mental process: The agent generates the beliefs, and the intention and action to cross the street	
3	Believing there is icecream across the street Having the intention to cross the street Generated action to cross the street		
		Downward transduction: *interaction from mental to material* Physical representation in the brain	
4			Physical representation of beliefs, intention and action initiation in the brain
		Downward action transduction: *interaction from mental to material* Action is to be executed	

TABLE 1 (*continued*)

State no.	Mental state	Process or interaction	Material state
5			Starting to cross the street Car approaching
		Material processes: Execution of the action and the event 'car appears' within the material world The crossing, approaching of the car and the collision process	
6			Agent at p2. Car at p2 Car has hit agent Body and brain damaged
		Upward transduction: *interaction from material to mental* The brain damage entails malfunctioning of short-term memory	
7	Absent short-term memory		
		Upward observation transduction: *interaction from material to mental* Observation of own position	
8	Observed agent is at p2		
		Mental process: processing new observations	
9	Belief agent is at p2		

processes can do its work in isolation, without being disturbed by physical events (except sensing). If within a model it is also taken into account that the agent's mind has a materialization in the form of a brain, the model for the interactions between mind and matter becomes much more complex. A generic knowledge level model is presented that is shown to be useful to model (and simulate) a variety of phenomena in which multiple mind–matter interactions occur.

Pylyshyn (1984) described the relations by the so-called transducers that connect aspects of the material world to symbolic representations and *vice versa*. Also from a practical agent modelling perspective, the division and relation between the agent and material world is not trivial. For example, a cup on a table can be considered part of the

TABLE 2
Example mind–matter trace for the E. coli case

State no.	Mental state	Process or interaction	Material state
0	Desiring to perform lactose import Desiring to perform glucose import		Lactose and glucose externally present
		Upward observation transduction: *interaction from material to mental* Observation of the external nutrients	
1	Observed glucose and lactose are externally present		
		Mental process: The agent generates beliefs, intention and action	
2	Believed glucose and lactose are externally present Having the intention to import glucose Generated action to import glucose		
		Downward action transduction: *interaction from mental to material* Action is to be executed	
3			Importing glucose Radiation event
		Material processes: Execution of the action and the radiation event within the material world	
4			Glucose import gene damaged
		Upward transduction: interaction from material to mental The gene damage entails the desire to become absent	

TABLE 2 (*continued*)

State no.	Mental state	Process or interaction	Material state
5	Absent glucose import desire		
		Mental process: Processing absence of glucose import desire	
6	Absent glucose import intention and action		
		Downward action transduction: interaction from mental to material Action is not to be executed	
7			Not importing glucose
		Material processes: Glucose disappears externally	
8			No glucose externally present
		Upward observation transduction: interaction from material to mental New observation	
9	Observed that no glucose is externally present		
		Mental process: Generating new intention and action	
10	Belief that no glucose is externally present Lactose import intention and action		
		Downward action transduction: interaction from mental to material	
11		Action is to be executed	
			Importing lactose

material world, but it is also convenient to consider material aspects of the agent as part of the world; for example a relation between the cup and a robot gripper that has picked up the cup then can be viewed as part of the structure of the material world. This perspective can be extended to a material world describing two agents shaking hands or even one agent, the left hand of which has gripped the right hand. These external material agent aspects (the agent's matter) can be modelled as different from the internal mental aspects of the agent such as its beliefs about the world, its intentions and plans, and its reasoning (the agent's mind). If it is also taken into account that the agent's mind has a materialization in the form of brain, the relations between mind and matter become more complex.

4.1. THE KNOWLEDGE REPRESENTATION HYPOTHESIS

Smith (1982, 1985) formulated the *knowledge representation hypothesis*. The essence of this hypothesis is the strict division between (a) the meaning of a *representation*, that can be attributed from outside, and (b) the *manipulation* of representation structures independent of their meaning, that is, it proceeds on the basis of form only.

In logic the knowledge representation hypothesis is the basis for formal systems. These systems formally define a language (in which formulae stand for the representations of knowledge), e.g. the language of predicate logic. The attribution of semantics is formalized by formal structures (called models, standing for world states the knowledge refers to); e.g. Tarski (1956), Dalen (1980), Chang and Keisler (1973) and Hodges (1993). For connections to reasoning systems, see e.g. Weyhrauch (1980) and Treur (1991). The manipulation of these syntactical structures is based on inference rules, such as modus ponens, conjunction introduction and others. These inference rules are defined in a generic manner: they do not depend on the meaning of the formulae on which they are applied.

Formal systems as defined in logic can be used to formalize cognitive representation systems and their (reference) relation with the material world they represent. However, there is a second type of relation between a cognitive system and the material world: the cognitive representations themselves are embodied in a material form in the brain.

4.2. TWO TYPES OF REPRESENTATIONS

Although it is not known in all details how the mental activities of a human agent take place, it is clear that the brain is an essential material aspect of it. Every thought is somehow physically embodied within the brain and every reasoning process is performed as a physical brain activity. This is a completely different relation between a mental or symbolic system and a material system that has nothing to do with the content of the symbolic representation (i.e. the material world aspects to which the representations refer), but only with the form in which the representation is materialized.

In this paper we interpret the materialization of representations as a second process of representation, to which again the knowledge representation hypothesis can be applied. For example, consider the concept of time. The symbolic representation noon can be represented in a material manner by a clock. A clock, a material piece of machinery, represents the symbol noon by the material configuration in which both hands of the clock point upward. Manipulations with these material representations take place according to physical laws that indeed (as demanded by the knowledge representation hypothesis) are independent of the content the representations refer to i.e. the movement of the hands of the clock just follow physical laws and are not affected in any manner by our attribution of semantics to the material configurations that occur.

Thus, following the knowledge representation hypothesis, it is not only possible to represent material aspects in a symbolic manner, but it is also possible to represent symbolic or mental aspects in a material manner. We distinguish the two types of representation as *material representation* vs. *mental* or *symbolic representation*. Dual representation relations are obtained (see Figure 4): material aspects of the world have a symbolic representation, and symbolic aspects have a material representation. Note that these relations are not related in a direct manner; e.g. they are not each other's inverse. Specific and bi-directional types of mind–matter interaction do occur frequently: observations in the material world affecting the information in the brain (sensing), mental processes leading to material actions affecting the world (acting), material processes in the world affecting the brain itself (e.g. causing brain damage) or mental processes affecting the material state of the body (e.g. causing psychosomatic diseases).

4.3. SIMULATION OF MATERIAL AND SYMBOLIC PROCESSES AND THEIR INTERACTIONS

The model developed in this paper simulates different types of mind–matter interaction. The material world in which agents live and think is depicted in Figure 5 at the right bottom. The cognitive symbolic system depicted at the right top represents the world and performs reasoning about the world (cf. Lindsay & Norman, 1977; Newell, 1980; Laird, Newell & Rosenbloom, 1987; Simon & Kaplan, 1989).

In order to make a model of the interacting material and symbolic processes that is executable on a computer system, a (formal) simulation model can be made. The simulation model is depicted on the left-hand side of Figure 5. It formalizes the following processes

(1) The material processes in the physical world.
(2) The symbolic processes in the cognitive system.
(3) The interaction between these two types of processes.

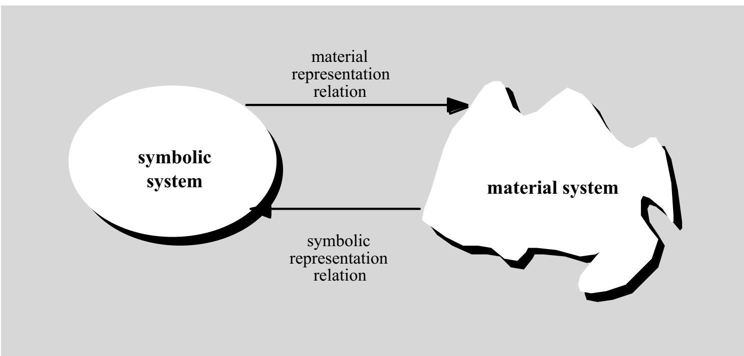

FIGURE 4. Dual representation relations.

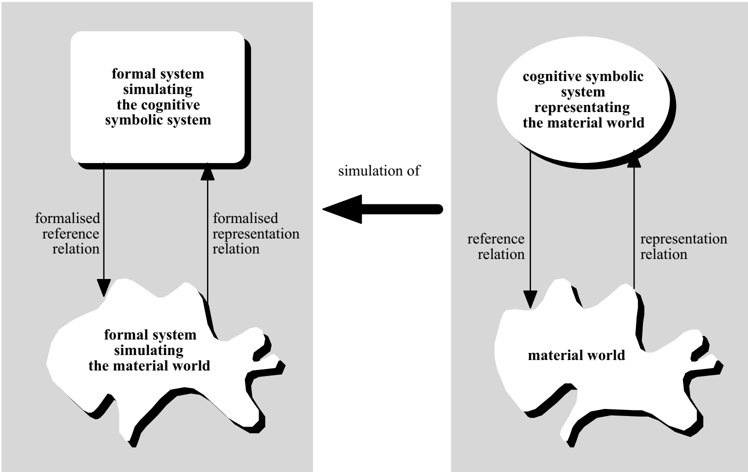

FIGURE 5. Simulating both the material world and the cognitive symbolic system representing it.

Note that a simulation does not pretend to have exactly the same behaviour as the original system: a rough approximation may be sufficient to obtain a specific insight into these processes.

To formalize material processes, a formal language is required in which physical laws can be expressed e.g. Newton or Leibniz's laws of mechanics. The quantitative mathematical language of calculus is often used for *quantitative modelling* of these processes. Quantitative mathematical models are often used in practice to simulate processes in the material world in order to predict them. Often these numerical simulation models are based on differential equations that specify how from a numerical description of one world state in the process, the next state can be calculated. The simulation is performed by repeating this step as often as is needed, thus generating a sequence of world states, or trace: M_0, M_1, M_2,.... If no information on precise numerical values is required but the aim is to describe qualitative phenomena, also qualitative modelling languages can be used to formalize physical processes, or languages that combine qualitative and quantitative elements. In the qualitative approach, often (this time qualitative) knowledge is specified that relates the next state of the world to the current state. A simulation of the world can be performed by applying this knowledge a number of times, thus also generating a sequence of world states. The recently developed approach of executable temporal logic (e.g. Barringer, Fisher, Gabbay, Owens & Reynolds, 1996) can be used for qualitative modelling. In this approach the world dynamics are specified in temporal rules of the form

$$A \,\&\, B \rightarrow C,$$

where A refers to the past of the process, B to the current state and C to the future states. In a simplified case A is left out and C refers to the next state. In this paper the world is simulated according to this simplified executable temporal logic approach.

5. Component-based design of agent systems

The model presented has been designed and specified using the component-based design method for multi-agent systems DESIRE (Design and Specification of Interacting Reasoning components); cf. Brazier, Jonker and Treur (1998), for a real-world case study, see Brazier, Dunin-Keplicz, Jennings and Treur (1995). In this section a brief overview of DESIRE is presented. The emphasis in DESIRE is on the conceptual and detailed design. The design of a multi-agent system in DESIRE is supported by graphical design tools within the DESIRE software environment. The software environment includes implementation generators with which (formal) design specifications can be translated into executable code of a prototype system. In DESIRE, a design consists of knowledge of the following three types: *process composition*, *knowledge composition* and the *relation* between process composition and knowledge composition. These three types of knowledge are discussed in more detail below.

5.1. PROCESS COMPOSITION

Process composition identifies the relevant processes at different levels of (process) abstraction, and describes how a process can be defined in terms of (is composed of) lower-level processes.

5.1.1. Identification of processes at different levels of abstraction. Processes can be described at different levels of abstraction: for example, the process of the multi-agent system as a whole, processes defined by individual agents and the external world, and processes defined by task-related components of individual agents. The identified processes are modelled as *components*. For each process the *input and output information types* are modelled. The identified levels of process abstraction are modelled as *abstraction/specialization relations* between components: components may be *composed* of other components or they may be *primitive*. Primitive components may be either reasoning components (i.e. based on a knowledge base) or components capable of performing tasks such as calculation, information retrieval and optimization. These levels of process abstraction provide process hiding at each level.

5.1.2. Composition of processes. The way in which processes at one level of abstraction are composed of processes at the adjacent lower abstraction level is called *composition*. This composition of processes is described by a specification of the possibilities for *information exchange* between processes (*static view* on the composition), and a specification of *task control knowledge* used to control processes and information exchange (*dynamic view* on the composition).

5.2. KNOWLEDGE COMPOSITION

Knowledge composition identifies the knowledge structures at different levels of (knowledge) abstraction, and describes how a knowledge structure can be defined in terms of lower-level knowledge structures. The knowledge abstraction levels may correspond to the process abstraction levels, but this is often not the case.

5.2.1. Identification of knowledge structures at different abstraction levels. The two main structures used as building blocks to model knowledge are *information types* and *knowledge bases*. Knowledge structures can be identified and described at different levels of abstraction. At higher levels details can be hidden. An *information type* defines an ontology (lexicon, vocabulary) to describe objects or terms, their sorts, and the relations or functions that can be defined on these objects. Information types can logically be represented in order-sorted predicate logic. A *knowledge base* defines a part of the knowledge that is used in one or more of the processes. Knowledge is represented by formulae in order-sorted predicate logic, which can be normalized by a standard transformation into rules.

5.2.2. Composition of knowledge structures. Information types can be composed of more specific information types, following the principle of compositionality discussed above. Similarly, knowledge bases can be composed of more specific knowledge bases. The compositional structure is based on the different levels of knowledge abstraction distinguished, and results in information and knowledge hiding.

5.3. RELATION BETWEEN PROCESS COMPOSITION AND KNOWLEDGE COMPOSITION

Each process in a process composition uses knowledge structures. Which knowledge structure is used for which process is defined by the relation between process composition and knowledge composition.

In Brazier, Jonker and Treur (2000) the component-based Generic Agent Model GAM is presented. This model covers the mental functioning of an agent for which interaction with its environment is restricted to sensing, acting and communication. Other types of mind–matter interaction are not covered in GAM.

6. A generic model for multiple mind–matter interaction

The generic model for multiple mind–matter interaction (see Figure 11) is presented in four different parts, described subsequently in Sections 6.1–6.4. In Section 6.5 the overall view is discussed.

6.1. THE MATERIAL WORLD AND ITS SYMBOLIC REPRESENTATION

In this section the material world and its symbolic representation, as well as the concept of transducers are discussed. The approach discussed in Section 1 will be applied. In Figure 6, the component material world simulates the actual material world. All changes with respect to physical aspects of objects take place within this component. The component symbolic representation of material world simulates the state of the symbolic representation of the material world over time. Both components and their interaction will be discussed in more detail in subsequent (sub)sections.

In order to reason about the material world and its behaviour, a symbolic representation of the material world is called for. In Figure 5, the component symbolic representation of material world specifies a simulation of such a representation.

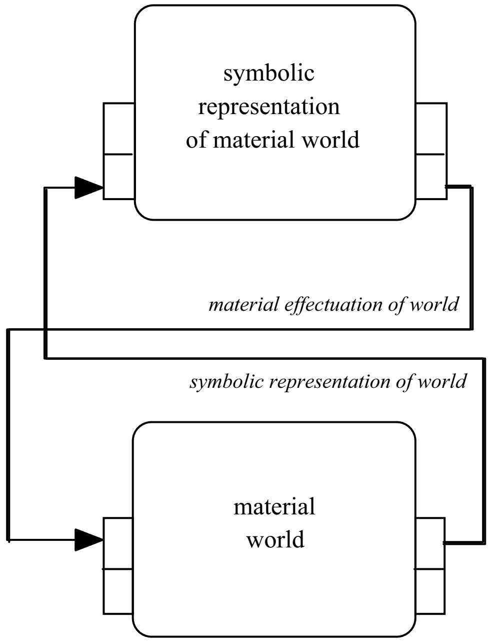

FIGURE 6. Transduction links between the material world and its symbolic representation.

6.1.1. Generic component descriptions. The vocabulary used for interaction with the component symbolic representation of material world is specified by the following information types both for input and output:

Input and output information type for symbolic representation of material world

sorts

WORLD_TERM;

meta-descriptions

material_world_it	WORLD_TERM;

functions

current_observation_result_of:	PROPERTY * SIGN * AGENT -> WORLD_TERM;

relations

to_be_performed_by :	ACTION * AGENT;
to_be_observed_by:	PROPERTY * AGENT;
just_acquired:	WORLD_TERM;

This information type introduces a new sort WORLD_TERM that is used in the construction of a meta-description of the information type material world it. In the meta-description all n-ary relations of the information type are transformed into n-ary functions into the sort WORLD_TERM. This construction allows, for example, the following atom:

just_acquired(current_observation_result_of(car_present, neg, agent))

Within the component symbolic representation of material world no knowledge is specified. The component in principle only models the maintenance of representation states. Also within this component updates are maintained, i.e. whenever an observation has been performed. Updates are specified by an information link from the component symbolic representation of material world to itself.

The vocabulary used for the inputs and outputs of the component material world consists of the following input and output information types:

Input information type for material world
  sorts
    ACTION, AGENT, PROPERTY;
  relations
    current_action_by: ACTION * AGENT;
    current_observation_by: PROPERTY * AGENT;
Output information type for material world
  sorts
    AGENT, PROPERTY, SIGN;
  relations
    current_observation_result_of: PROPERTY * SIGN * AGENT;

In addition, depending on the application of the model, specific information types are added.

6.1.2. Interaction between material and symbolic representation of the world. As discussed in Section 1, there are two issues in using a symbolic representation of the material world. The first is how changes in the material world become reflected in the symbolic representation (upward transduction). The second is how changes in the symbolic representation of the world affect the material world itself (downward transduction). In Figure 6, the simulations of transducers are modelled within the framework DESIRE as links between the output and input states of the components material world and symbolic representation of material world. The links that model transducers are called *transduction links* (and depicted in italics). The downward transducer is modelled by the transduction link *material effectuation of world*, the upward by the transduction link *symbolic representation of world*. The downward link transfers actions that are to be performed to the component material world. Moreover, observations can be made. The results of observations are transferred to the component symbolic representation of material world, by way of the transduction link *symbolic representation of world*, during which a symbolic representation of the observation results is made that can be processed by the receiving component. In Figure 6 each component has a levelled interface (denoted by the rectangles on the side of the components). The transduction link *symbolic representation of world* transfers epistemic meta-level information on the material world (e.g. expressed by the truth of the atom true(current_observation_result_of(car_present, pos, agent))) to object-level information that can be used by the component symbolic representation of material world (expressed by the truth value of the atom just_acquired (current_observation_result_of(car_present, pos, agent))): the atom links

```
(      true(current_observation_result_of(P : PROPERTY, S : SIGN, agent)),
       just_acquired(current_observation_result_of(P : PROPERTY, S : SIGN, agent))
)      : <<true,true>>;

(      true(current_observation_result_of(P : PROPERTY, S : SIGN, agent)),
       just_acquired(current_observation_result_of(P : PROPERTY, neg, agent))
)      : <<false,true>> ;
```

The transduction link *material_effectuation_of_world* links information on actions to be executed of the component symbolic_representation_of_material_world, to meta-level information on the material world: the atom links

```
(      to_be_performed_by(A : ACTION, agent),
       assumption(current_action_by(A : ACTION, agent), pos)
)      : << true,true> <false,false>, <unknown,unknown>> ;

(      to_be_observed_by(P : PROPERTY, agent),
       assumption(current_observation_by(P : PROPERTY, agent), pos)
)      : <<true,true>, <false,false>, <unknown,unknown>> ;
```

In this example, the truth value combinations <false,false> and <unknown,unknown> ensure that previous actions are retracted, so that actions will not be performed *ad infinitum*.

6.2. AN AGENT'S BEHAVIOUR IN INTERACTION WITH THE MATERIAL WORLD

As discussed in Section 4 the downward transduction link is needed for the actual execution of actions. However, the component symbolic representation of material world is not modelled as a component in which decisions are made on which observation or action is to be performed and when (pro-active behaviour). Such mental decision processes are modelled in the component agent, see Figure 7.

6.2.1. Generic description of agent. The component agent models the cognitive symbolic reasoning system of an agent as a logical system. The agent can determine observations and actions to be performed. The vocabulary used for interaction with the component agent is specified by the following generic input and output information types.

Input information type for agent
  sorts
    WORLD_TERM ;
  meta-descriptions
    material_world_it :   WORLD_TERM ;
  functions

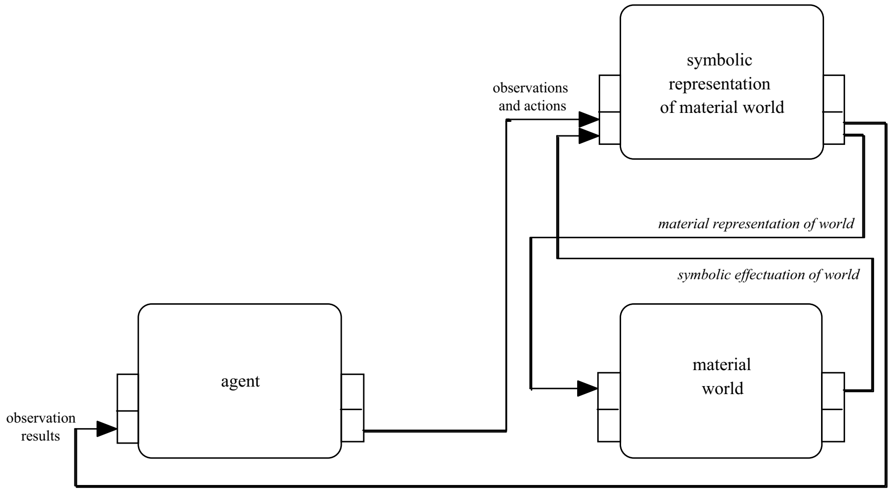

FIGURE 7. Transduction and symbolic links connecting agent and material world.

current_observation_result : PROPERTY * SIGN -> WORLD_TERM ;
relations
just_acquired: WORLD_TERM ;

Output information type for agent
sorts
ACTION, PROPERTY ;
relations
to_be_performed: ACTION ;
to_be_observed: PROPERTY ;

Note again the meta-description construct within this information type.

6.2.2. Symbolic links for interaction with agent. The symbols representing the decisions to perform observations and actions are linked to the symbolic system modelled by the component symbolic representation of material world. All connections between symbolic systems are called *symbolic links*. Symbolic links are modelled as information links within the framework DESIRE. The symbolic link that transfers the symbolic representations of observations and actions that are to be performed is called observations and actions. This link connects the object level of the output interface of the component agent with the object-level input interface of the component symbolic representation of material world: atom links

(to_be_observed(P : PROPERTY)
to_be_observed_by(P : PROPERTY, agent)

)   : <<true,true>,<false,false>,<unknown, unknown>> ;

(to_be_performed(A : ACTION)
         to_be_performed_by(A : ACTION, agent)
)   : <<true,true> <false,false> <unknown,unknown>> ;

The results of observations performed within material world are transferred to the component agent through the transduction link *symbolic representation world* (see Section 6.1) and the symbolic link observation results that connects the component symbolic representation of material world to the component agent: atom links

(just_acquired(current_observation_result_of(X : PROPERTY, S : SIGN, agent)),
         just_acquired(current_observation_result(X : PROPERTY, S : SIGN))
)   : <<true,true>, <false,false>> ;

6.3. AN AGENT AND ITS MATERIAL REPRESENTATION

In Figure 8, the cognitive symbolic system of the agent is modelled by the component agent described in the previous sub-section. The component material representation of

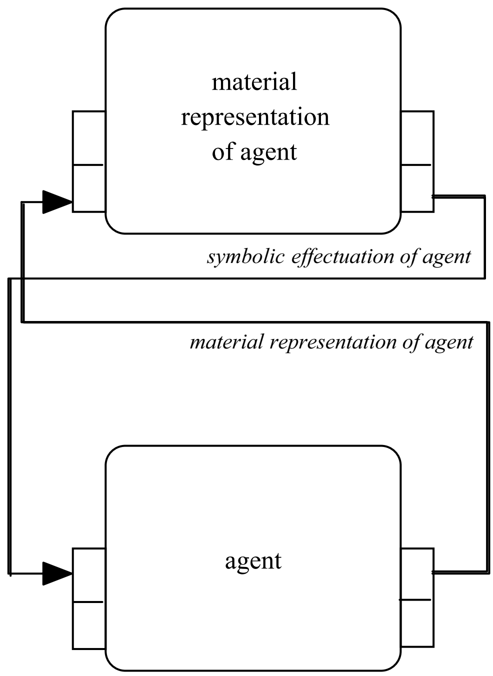

FIGURE 8. Transduction links between the agent and its material representation.

agent models the material representation of (the symbolic system of) the agent. As discussed in Section 1, the relation between the agent and its material representation is modelled in a manner similar to the manner in which the relation between the material world and its symbolic representation is modelled. An upward transducer defines how symbolic aspects of the agent are represented in a material form, a downward transducer defines how properties of the material world affect the processes of the symbolic system within the agent.

6.3.1. Generic description of the material representation of the agent. The vocabulary used for interaction with the component material representation of agent is specified by the following input and output information types.

Input information type for material representation of agent

sorts

AGENT_ATOM, SIGN ;

meta-descriptions

symbolic_agent_it: AGENT_ATOM ;

relations

to_be_stored: AGENT_ATOM*SIGN ;

For simplicity in this paper it is assumed that there exist functions that relate information in memory to locations within the brain, i.e. positions:

$$\text{position(I : INFORMATION\_OBJECT, B : BRAIN\_LOCATION)}$$

The information types of agent are used in a meta-description construct, such that the relations of that information type can be used as functions into the sort AGENT_ATOM. The sort AGENT_ATOM is a sub-sort of the sort PROPERTY. Therefore, all relations that have PROPERTY as an argument can be applied to the new terms.

The simple model for memory used in this paper has a short-term memory and a long-term memory. To model this distinction, the sort BRAIN_LOCATION has two sub-sorts: STM_LOCATION and LTM_LOCATION. Given the atom of the agent (a term of the sort AGENT_ATOM) and a time point (a term of the sort TIME), the function stm_location relates information to a position within the short-term memory, whereas ltm_location relates information to a position within the long-term memory. The time point used by the function is the moment in time that the information is stored into the memory. An information object is specified as

$$\text{information\_object(A : AGENT\_ATOM, S : SIGN),}$$

where the sort AGENT_ATOM contains objects that refer to atoms of the agent, e.g. observed_at(car_present, neg, t1). The current status of the memory is modelled by atoms of the form

$$\text{currently(position(information\_object(A : AGENT\_ATOM, S : SIGN),}$$
$$\text{B : BRAIN\_LOCATION), pos)}$$

To specify this for the output the following information type is used:

```
Output information type for material representation of agent
  sorts
    AGENT_ATOM, SIGN ;
  meta-descriptions
    symbolic_agent_it:            AGENT_ATOM ;
  functions
    information_object:           AGENT_ATOM*SIGN -> OBJECT
    position:                     OBJECT*POSITION -> PROPERTY
  relations
    currently:                    PROPERTY*SIGN ;
```

6.3.2. Transduction links between agent and its material representation. The information maintained by the agent is built of atoms with an explicit reference to their truth-value in the form of a sign. The atom is transformed into a term by the transition from the agent to its material representation. For example, the atom observed_at(car_present, neg, t1) that can be used within the component agent is represented by a *term* within the component material representation of agent. If the atom is true within agent, the sign pos is to be added within material representation of agent, if the atom was false, the sign neg is to be added. If the atom has the truth-value unknown it is not stored in material representation of agent. If the agent, by reasoning, makes cognitive changes in its beliefs, desires, intentions or knowledge, the material representations of these changes are materialized in the brain. This process of upward transduction is modelled by the transduction link *material representation of agent*. Example atom links are

```
(     true(A : IIOA),
      to_be_stored(A : AGENT_ATOM, pos)
)     : <<true,true>> ;

(     false(A : IIOA),
      to_be_stored(A : AGENT_ATOM, neg)
)     : <<true,true>> ;
```

An example of an instantiated atom link of *representation info from OPC* is

```
(     true(observed_at(car_present, neg, t1)),
      to_be_stored(observed_at(car_present, neg, t1), pos)
)     : <<true,true>> ;
```

An example of an instantiated atom link of *material representation of agent* is

```
(     to_be_stored(observed_at(car_present, neg, t1), pos),
      to_be_stored(observed_at(car_present, neg, t1), pos)
)     : <<true,true>> ;
```

If a physical change within the component material representation of agent occurs, the symbolic interpretation of the changed information is linked to the component agent by the downward transduction process, modelled by the transduction links *symbolic effectuation of agent*, *effectuation info to OPC* and *effectuation info to MWI*. The atom links of the transduction link *effectuation info to OPC* are specified as follows:

(currently(position(information_object(A : AGENT_ATOM, S : SIGN), B : STM_LOCATION), pos),
assumption(A : AGENT_ATOM, S : SIGN)
): < <true,true>, <false,false>, <unknown,unknown> > ;

By these transduction links object-level information from the component material representation of agent is transferred to meta-level information within the component agent, which defines the current information state of the agent.

6.4. THE MATERIAL WORLD'S PHYSICAL BEHAVIOUR IN INTERACTION WITH THE AGENT

The material representation of the agent is a part of the material world. Therefore, the component material representation of agent is modelled as a simple component for passing on information. The material links connecting these two components (see Figure 9), update material world and update material representation of agent, are simple identity links, i.e. they only transfer information, they do not translate it. For example, the material link update material representation of agent links atoms to themselves:

(at_time(position(I : INFORMATION_OBJECT, B : BRAIN_LOCATION), S : SIGN, T : TIME) ,
at_time(position(I : INFORMATION_OBJECT, B : BRAIN_LOCATION), S : SIGN, T : TIME)
) : < <true,true>, <unknown,unknown>, <false,false> > ;

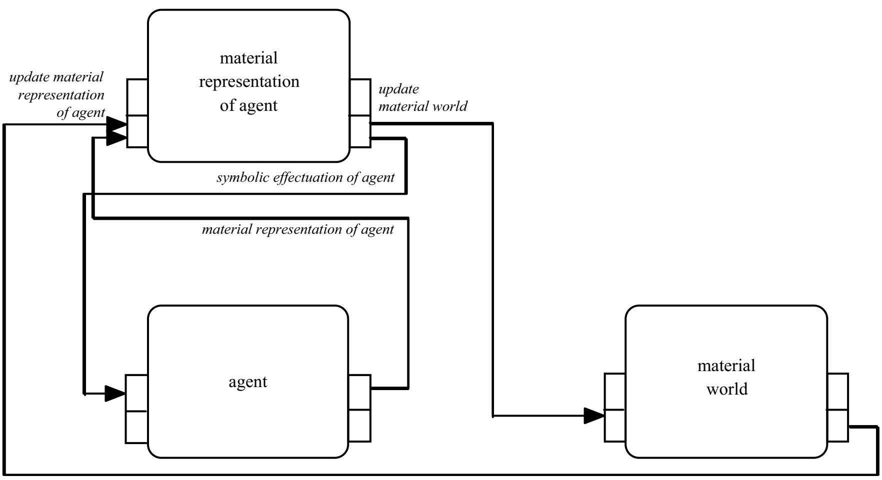

FIGURE 9. Transduction and material links connecting material world and agent.

6.5. THE COMPLETE MODEL AND ITS TRACE SEMANTICS

As can be seen from Figures 6–9 it is possible to create a symbolic representation of a material system and to create a material representation of a symbolic system. In Figure 10, all components and all information links (transduction, symbolic and material links) of the top level of the complete model are presented. Together, they sketch two connections between the agent and the material world. The connection between material representations and symbolic representations is made by transduction links, between symbolic representations by symbolic links and between material representations by material links.

The model as described only makes use of the types of transitions discussed in Section 3. The component agent (together with symbolic representation of material world) makes symbolic (reasoning) state transitions, the component material_world (together with material representation of agent) makes material world state transitions. Based on the material representation state of the agent, the link symbolic effectuation of agent makes downward transduction transitions for the agent, changing the symbolic representation of the agent. Based on the symbolic state of the agent, the link material representation of agent makes upward transduction transitions of the agent, changing the agent's material representation. Analogously, based on the symbolic representation state of the world, the link material representation of world makes downward transduction transitions for the world, changing the material world state. Moreover, based on the material world state, the link symbolic effectuation of world makes upward transduction transitions of the world, changing the world's symbolic representation.

7. Application of the multiple mind–matter interaction model by refinement

The generic model for multiple mind–matter interaction described in Section 6 has been applied to the two case studies introduced in Section 2. Application of this generic model entails refining the model by specialization (i.e. composing components from new, more fine-grained sub-components) and by instantiating application-specific information types and knowledge bases to specify the functionality of components.

7.1. REFINEMENT OF THE MODEL TO THE ICECREAM EXAMPLE

7.1.1. Instantiation within the material world. As discussed in Section 4, the material world is simulated by a specification in terms of executable temporal rules. The vocabulary within the component material world in which these temporal rules are expressed is defined by the following information types.

Information types for material world

sorts

ACTION, AGENT, AGENT_PROPERTY, EVENT, OBJECT, POSITION, PROPERTY, SIGN, TIME ;

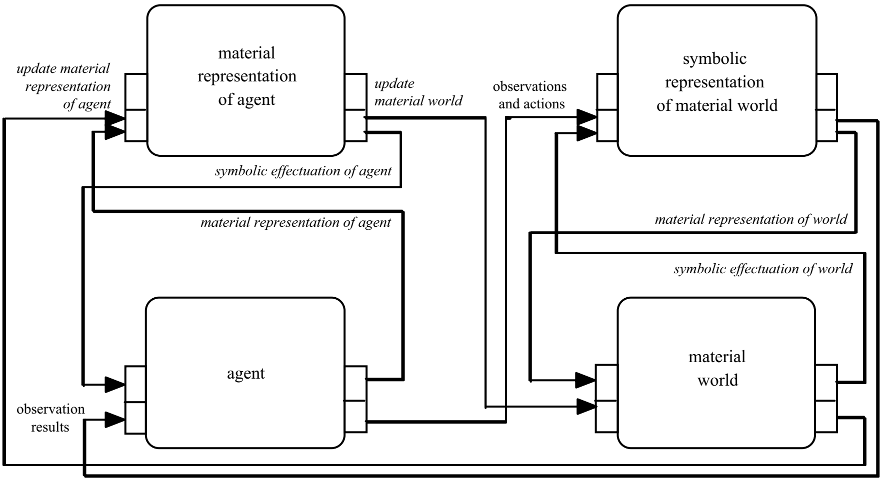

FIGURE 10. Transduction, symbolic and material links connecting agent and material world.

sub-sorts	
ACTION :	EVENT ;
AGENT :	OBJECT ;
AGENT_PROPERTY :	PROPERTY ;
objects	
agent :	AGENT ;
neg, pos :	SIGN ;
t0, t1, t2, t3 :	TIME ;
functions	
position:	OBJECT * POSITION -> PROPERTY ;
relations	
at_time:	PROPERTY * SIGN * TIME ;
current_time:	TIME ;
currently :	PROPERTY * SIGN ;
effect :	EVENT * PROPERTY * SIGN ;
event_after :	EVENT * TIME ;
event_to_happen :	EVENT ;
next :	PROPERTY * SIGN ;
next_time_point :	TIME ;
precedes :	TIME * TIME
objects	
car_to_appear:	EVENT ;
car, icecream, supermarket:	OBJECT ;
p1, p2, p3:	POSITION ;
car_present:	PROPERTY ;

functions

close_by_for:	OBJECT * AGENT -> PROPERTY ;
goto:	POSITION -> ACTION ;
has_hit:	OBJECT * OBJECT -> PROPERTY ;
next_on_path:	POSITION * POSITION * POSITION -> PROPERTY ;
has_sign:	OBJECT * OBJECT -> PROPERTY ;

sorts

AGENT_ATOM, BRAIN_LOCATION, INFORMATION_OBJECT, LTM_LOCATION, STM_LOCATION;

sub-sorts

AGENT_ATOM :	AGENT_PROPERTY ;
BRAIN_LOCATION :	POSITION ;
LTM_LOCATION :	BRAIN_LOCATION ;
STM_LOCATION :	BRAIN_LOCATION ;
INFORMATION_OBJECT:	OBJECT ;

functions

contents_of_stm_to_ltm:	INFORMATION_OBJECT * STM_LOCATION -> EVENT ;
has_amnesia:	AGENT -> AGENT_PROPERTY ;
information_object:	AGENT_ATOM * SIGN -> INFORMATION_OBJECT ;
ltm_location:	INFORMATION_OBJECT * TIME -> LTM_LOCATION ;
recovered:	AGENT -> AGENT_PROPERTY ;
recovering:	AGENT -> EVENT ;
stm_location:	INFORMATION_OBJECT * TIME -> STM_LOCATION ;
to_be_stored:	AGENT_ATOM * SIGN ;

The (temporal) knowledge simulating the processes within the material world is specified as follows:

/ domain dependent knowledge */*
at_time(position(supermarket, p3), pos, T: TIME) ;
at_time(has_sign(supermarket, icecream), pos, T: TIME) ;
at_time(position(agent, p1), pos, t1) ;
at_time(car_present, neg, t1) ;
effect(car_to_appear, position(car, p1), neg) ;
effect(car_to_appear, position(car, p2), pos) ;
effect(goto(P : POSITION), position(agent, P : POSITION), pos) ;
effect(recovering(X : OBJECT), has_amnesia(X : OBJECT), neg) ;
event_after(car_to_appear, t1) ;
precedes(t0, t1) ;
precedes(t1, t2) ;
precedes(t2, t3) ;

if at_time(position(A : AGENT, p1), pos, T: TIME)
and at_time(position(O:OBJECT,p3), pos, T:TIME)
then at_time(close_by_for(O : OBJECT, A : AGENT), pos ,T : TIME) ;

if current_observation_by(P : PROPERTY, A : AGENT)
and current_time(T : TIME)
and at_time(P : PROPERTY, S : SIGN, T : TIME)
then current_observation_result_of(P : PROPERTY, S : SIGN, A : AGENT) ;

/ knowledge about time and events */*

if at_time(position(X : OBJECT, P : POSITION), pos, T1 : TIME)
and at_time(position(Y : OBJECT, Q : POSITION), pos, T1 : TIME)
and not equal(X : OBJECT, Y : OBJECT)
and not equal(P : POSITION, Q : POSITION)
and precedes(T1 : TIME, T2 : TIME)
and at_time(position(X : OBJECT, R : POSITION), pos, T2 : TIME)
and at_time(position(Y : OBJECT, R : POSITION), pos, T2 : TIME)
then at_time(has_hit(X : OBJECT, Y : OBJECT), pos, T2 : TIME) ;

if at_time(has_hit(X : OBJECT, agent), pos, T : TIME)
then at_time(has_amnesia(agent), pos, T : TIME) ;

if current_action_by(A : ACTION, X : AGENT)
and effect(A : ACTION, P : PROPERTY, S : SIGN)
then next(P : PROPERTY, S : SIGN) ;

if event_to_happen(E : EVENT)
and effect(E : EVENT, P : PROPERTY, S : SIGN)
then next(P : PROPERTY, S : SIGN) ;

if currently(has_amnesia(X : OBJECT), pos)
then event_to_happen(recovering(X : OBJECT)) ;

if current_time(T2 : TIME)
and precedes(T1 : TIME, T2 : TIME)
and at_time(position(I : INFORMATION_OBJECT, B: STM_LOCATION), pos, T1 : TIME)
then event_to_happen(contents_of_stm_to_ltm(I : INFORMATION_OBJECT, B: STM_LOCATION)) ;

if current_time(T1 : TIME)
and precedes(T1 : TIME, T2 : TIME)
then effect(contents_of_stm_to_ltm(I : INFORMATION_OBJECT, B: STM_LOCATION), position(I : INFORMATION_OBJECT, ltm_location(I : INFORMATION_OBJECT, T2 : TIME)), pos) ;

if currently(has_amnesia(X : AGENT), neg)
and current_time(T2 : TIME)
and precedes(T1 : TIME, T2 : TIME)
and not event_to_happen(contents_of_stm_to_ltm(I : INFORMATION_OBJECT, B: STM_LOCATION))
and at_time(position(I : INFORMATION_OBJECT, B: STM_LOCATION), pos, T1 : TIME)
then at_time(position(I : INFORMATION_OBJECT, B: STM_LOCATION), pos, T2 : TIME) ;

```
if      current_time(T2 : TIME)
        and precedes(T1 : TIME, T2 : TIME)
        and at_time(I : INFORMATION_OBJECT, B: LTM_LOCATION), pos, T1 : TIME)
then    at_time(I : INFORMATION_OBJECT, B: LTM_LOCATION), pos, T2 : TIME) ;

if      current_time(T : TIME)
        and at_time(P : PROPERTY, S : SIGN, T : TIME)
then    currently(P : PROPERTY, S : SIGN) ;

if      event_after(E : EVENT, T : TIME )
        and current_time(T : TIME)
        then event_to_happen(E : EVENT ) ;
if      not equal(P : POSITION, Q : POSITION)
then    effect(goto(P : POSITION), position(agent, Q : POSITION), neg) ;

if      current_time(T1 : TIME)
        and precedes(T1 : TIME, T2 : TIME)
then    next_time_point(T2 : TIME) ;

if      next_time_point(T2 : TIME)
        and next(X : PROPERTY, S : SIGN)
then    at_time(X : PROPERTY, S : SIGN, T2 : TIME) ;
```

To execute the temporal rules specified above, updates are required from the current time point to the next time point. These updates are specified by an information link from the component material world to itself.

Within the component material world a simple model for memory is specified. The component material representation of agent only maintains a state, from and to which information is transferred to and from the component material world. The only exception is the following knowledge base rule that combines the information to be stored and the current time point and determines the actual storage of the information as a physical property:

```
if      current_time(T : TIME)
        and to_be_stored(A : AGENT_ATOM, S : SIGN)
then    at_time(position(information_object(A : AGENT_ATOM, S : SIGN), stm_location(A :
        INFORMATION_OBJECT (A: AGENT_ATOM, S : SIGN), T : TIME), pos, T : TIME) ;
```

7.1.2. Specialization and instantiation within agent.

Information type for agent

```
  sorts
                                        WORLD_TERM ;
  meta-descriptions
    material_world_it :                 WORLD_TERM ;
```

functions	
current_observation_result :	PROPERTY * SIGN -> WORLD_TERM ;
close_by :	OBJECT -> PROPERTY ;
own_position :	POSITION ;
visiting :	AGENT -> WORLD_TERM ;
relations	
belief:	WORLD_TERM ;
current_belief:	PROPERTY * SIGN ;
desire:	OBJECT ;
intention:	WORLD_TERM ;
most_recent_observation:	PROPERTY * SIGN ;
observed:	PROPERTY ;
observed_at:	PROPERTY * SIGN * TIME ;
possible_observation:	PROPERTY ;
to_be_performed:	ACTION ;
to_be_observed:	PROPERTY ;

The agent is modelled as a composed component consisting of two sub-components, own process control and maintain world information, see Figure 11. The reasoning about its intentions, desires and plans is performed within the component own process control. Its knowledge about the world, obtained by observations, is maintained within the component maintain world information.

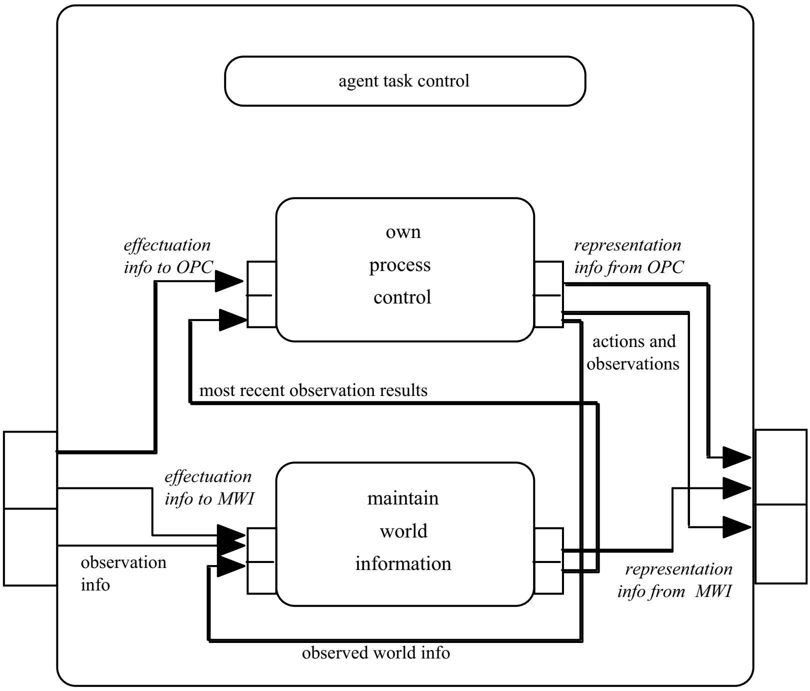

FIGURE 11. Links within the agent.

The component own process control contains the following knowledge:

```
desire(icecream) ;
to_be_observed(own_position(P : POSITION));
to_be_observed(car_present);

if      desire(G : OBJECT)
then    possible_observation(position(S : OBJECT, P : POSITION))
        and possible_observation(close_by(S : OBJECT))
        and possible_observation(has_sign(S : OBJECT, G : OBJECT));

if      possible_observation(P : PROPERTY)
        and not observed(P : PROPERTY)
then    to_be_observed(P : PROPERTY);

if      current_belief(has_sign(S : OBJECT, G : OBJECT), pos)
        and desire(G : OBJECT)
        and current_belief(close_by(S : OBJECT), pos)
then    intention(visiting(S : OBJECT));

if      intention(visiting(S : OBJECT))
         and current_belief(position(S : OBJECT, P : POSITION), pos)
         and current_belief(own_position(Q : POSITION), pos)
         and current_belief(car_present, neg)
         and current_belief(next_on_path(R : POSITION, Q : POSITION, P : POSITION), pos)
then    to_be_performed(goto(R : POSITION));

if      current_time(T : TIME)
         and just_acquired(current_observation_result(P : PROPERTY, S : SIGN))
then    observed_at(P : PROPERTY, S : SIGN, T : TIME);
```

The link observed world info transfers the just acquired knowledge about the world from own process control to maintain world information. The agent obtains this knowledge by observations. The link updates the truth-values of the atom most_recent_observation ensuring that the atom indeed reflects the most recent information about the world. The link most recent observation results determines the beliefs that are to be held by the agent (within its component own process control).

7.2. REFINEMENT OF THE MULTIPLE MIND–MATTER MODEL TO THE *E.COLI* EXAMPLE

The generic mind–matter interaction model can be refined to an application simulating the dynamics of the *E. coli* case by instantiating the information types and knowledge involved. No further sub-components are needed.

7.2.1. Instantiation within the material world. Instantiation of the material world within the model to the *E. coli* example involves information types to model the physical effects of a radiation source on the bacterium's DNA. Typically this effect would be statistical in nature. To obtain an explanation of the dynamic pattern discussed in Section 2.2, the simplification is made that the radiation causes the gene for glucose import to be damaged while leaving the rest unimpaired. This is modelled by

```
effect(radiation, impaired_glucose_import_gene, pos)

if      current_time(t3)
then event_to_happen(radiation));

if      event_to_happen(E : EVENT )
        and effect(E : EVENT, P : PROPERTY, S : SIGN)
then    next(P : PROPERTY, S : SIGN) ;
```

The link from material to mental description of the agent this time includes the atom link

```
  (     false(impaired_glucose_import_gene),
        desire(glucose_import)
  ):    <true, true>, <false, false>
```

7.2.2. Instantiation within the agent. Within the agent knowledge is specified that allows it to generate appropriate intentions and actions under certain environmental conditions; see also Jonker *et al*. (2001, 2002).

```
if        desire(glucose_import)
then      intention(glucose_import)

 if       intention(glucose_import)
   and    current_belief(glucose_externally_present)
 then     to_be_performed(glucose_import)

 if       desire(lactose_import)
   and    current_belief(not glucose_externally_present)
   and    current_belief(lactose_externally_present)
 then     intention(lactose_import)

 if       intention(lactose_import)
   and    current_belief(not glucose_externally_present)
   and    current_belief(lactose_externally_present)
 then     to_be_performed(lactose_import)
```

Here the beliefs are created from observation results in a manner similar to the model for the icecream example.

8. Simulation traces for the example mind–matter interaction patterns

In this section first in Table 3 it is shown how the course of events in the ice-cream example introduced in Section 2 is simulated as a reactive pattern using the model introduced in the previous sections. The simulation trace is started at the moment that the agent is in position p1 and has observed that a supermarket where icecream is sold is

TABLE 3
Simulation trace for the icecream example

No.	Mental state	Process component or interaction links	Material state
0	desire(icecream)		position(supermarket, p3) has_sign(supermarket, icecream)
		symbolic effectuation of world; observation results	
1	observed (at_position (supermarket, p3)) observed(has_sign (supermarket, icecream))		
		material representation of agent; update material world	
2			STM properties for the two observed facts: position(information_ object(a1, pos), b1) position(information_ object(a2, pos), b2) with a1, resp. a2 representing the atoms observed(at_position (supermarket, p3)) observed(has_sign (supermarket, icecream)) and b1, b2 STM locations
		agent	
3	current_belief (at_position (supermarket, p3)) current_belief (has_sign(supermarket, icecream)) intention(visiting_ supermarket) to_be_performed(go- _performed(goto(p2))		
		material representation of agent; update material world	

TABLE 3 (*continued*)

No.	Mental state	Process component or interaction links	Material state
4			STM properties for the two beliefs, the intention, and the action initiation: position(information_object(a3, pos), b3) position(information_object(a4, pos), b4) position(information_object(a5, pos), b5) position(information_object(a6, pos), b6) with a3, …, a6 representing the atoms current_belief(at_position(supermarket, p3)) current_belief(has_sign(supermarket, icecream)) intention(visiting_supermarket) to_be_performed(go-_performed(goto(p2)) and b3, …, b6 STM locations
		observations and actions; *material representation of world*	
5			current_action_by(goto(p2), agent) event_to_happen(car_to_appear)
		material world	
6			position(agent, p2) position(car, p2) has_hit(car, agent) has_amnesia(agent) not position(information_object(a1, pos), b1) not position(information_object(a2, pos), b2) not position(information_object(a3, pos), b3)

TABLE 3 (*continued*)

No.	Mental state	Process component or interaction links	Material state
			not position(information_object(a4, pos), b4) not position(information_object(a5, pos), b5) not position(information_object(a6, pos), b6)
		update material representation of agent; symbolic effectuation of agent	
7	not current_belief (at_position (supermarket, p3)) not current_belief (has_sign(supermarket, icecream)) not intention (visiting_supermarket) not to_be_performed (goto(p2))		
		symbolic effectuation of world; observation results	
8	observed(at_position (agent, p2))		
		agent	
9	current_belief (at_position(agent, p2))		

at position p3, and that a path from p1 to p3 is available with p2 as next position. Moreover, the agent has observed that no car was present. These observations were made using the transduction links material representation of world (to execute the initiated observation) and symbolic effectuation of world (to acquire the observation results) between the material world and its symbolic representation, and the symbolic links observations and actions (to initiate the observation) and observation results (to pass the observation results). As a result the observation information is available within the agent (as current beliefs). The trace is started at time point t1. For a summary of the informal trace, see Table 1. In a similar manner, Table 4 depicts a simulation trace for the *E. coli* case; for an informal trace, see Table 2.

TABLE 4
Simulation trace for the E. coli example

No.	Mental state	Process component or interaction links	Material state
0	desire(glucose_import) desire(lactose_import)		glucose_externally_pre-_present lactose_externally_present not impaired_glucose_import_gene not impaired_lactose_import_gene
		symbolic effectuation of world; observation results	
1	observed(glucose_externally_present) observed(lactose_externally_present)		
		agent	
2	current_belief (glucose_externally_present) current_belief (lactose_externally_present) intention(glucose_import) to_be_performed (glucose_import)		
		observations and actions; *material representation of world*	
3			current_action_by (glucose_import, agent) event_to_happen (radiation)
		material world	
4			impaired_glucose_import_gene
		update material representation of agent; *symbolic effectuation of agent*	
5	not desire(glucose import)		
		agent	

TABLE 4 (*continued*)

No.	Mental state	Process component or interaction links	Material state
6	not intention(glucose import) not to_be_performed (glucose import)		
		observations and actions; *material representation of world*	
7			not current_action_by (glucose_import, agent)
		material world	
8			not glucose_externally_present
		symbolic effectuation of world; observation results	
9	observed(not glucose_externally_present)		
		agent	
10	current_belief(not glucose_externally_present) not current_belief (glucose_externally_present) intention(lactose_import) to_be_performed(lactose_import)		
		observations and actions; *material representation of world*	
11			current_action_by (lactose_import, agent)

9. Modelling some other patterns of multiple mind–matter interaction

In the previous sections the example course of events was simulated as a reactive pattern through Figure 10 from the lower left-hand side component (agent) to the upper right-hand side component (symbolic representation of material world) to the lower right-hand side component (material world) to the higher left-hand side component (material representation of agent) to the lower left-hand side component (agent). Also for various other types of interaction between symbolic systems and material systems such patterns can be identified. In this section a number of examples are discussed.

9.1. DRUG USE

Using the model introduced in this paper the process of taking a (narcotic) drug can be simulated as follows (see Figure 10):

Decision of the agent to take the drug
Reasoning within the component agent; deriving conclusion to_be_performed (take_drug).

Transfer the action to the material world
By the symbolic link observations and actions to the component symbolic representation of material world and by the downward transduction link *material effectuation of world* to the component material world.

Execution of the action take drug within the material world
Determination of the effect active_brain of the action take_drug.

Transfer the effects of take drug to the agent
By the material link update material representation of agent and the downward transduction link *symbolic effectuation of agent* to the component agent.

Execution of the agent with drug effect

9.2. AGENTS PLANNING AND EXECUTING BIRTH AND DEATH

Using the model introduced in this paper the process of creating a new child agent by a rational agent can be simulated by a similar pattern in Figure 10 (see also Brazier, Jonker, Treur & Wijngaards, 2001):

Decision of the agent to create a child agent
Reasoning within the component agent; deriving conclusion to_be_performed (create_child).

Transfer the action to the material world
By the symbolic link observations and actions to the component symbolic representation of material world and by the downward transduction link *material effectuation of world* to the component material world.

Execution of the action create child within the material world
Determination of the effect of the action create_child.

Transfer the effects of to create child to the agent
By the material link update material representation of agent and the downward transduction link *symbolic effectuation of agent* to the component agent; this link modifies the component agent by replacing it by two similar components

Execution of the agent and its child agent

In a similar manner a rational action to kill an agent can be modelled.

9.3. PSYCHOSOMATIC DISEASES

For psychosomatic diseases the pattern in Figure 10 proceeds in a different direction: from the lower left-hand side component to the upper left-hand side component to the lower right-hand side component. For example, a heart attack induced by psychological factors can be modelled as follows:

The agent reasons about highly stress-provoking information
Stressful reasoning within the component agent.

Transfer of the stress to the material representation of the agent
By the upward transduction link *material representation of agent* to the component material representation of agent (to the property over_active_brain) and by the material link update material world to the component material world.

Execution of the material world
Determination of the effect of over_active_brain on heart functioning.

9.4. MODELLING APPLICATIONS OF DIRECT BRAIN–COMPUTER INTERFACING

Recently, substantial progress has been made in the area of direct brain–computer interfacing; cf. Birbaumer *et al.* (1999, 2000) and Levine *et al.* (1999, 2000). The significance of this progress leads to high expectations for the development of applications of direct brain–computer interfacing in the area of Rehabilitation Engineering in the near future (e.g. Robinson, 2000). For modelling such applications involving direct brain–computer interfacing using the multiple mind–matter interaction model presented in this paper, it is assumed that two agents are involved: the human agent, and the supporting computer agent, for example related to a wheel chair. For this situation the pattern through Figure 10 proceeds as follows.

(1) For the human agent from the lower left-hand side component to the upper left-hand side component to the lower right-hand side component.
(2) For the computer agent from the lower right-hand side component via the upper right-hand side component to the lower left-hand side component.
(3) For the computer agent from the lower left-hand side component via the upper right-hand side component back to the lower right-hand side component.
(4) For the human agent from the lower right-hand side component to the upper left-hand side component, and from there to the lower left-hand side component.

In more detail, the bi-agent process runs according to a double loop in the following manner:

The human agent reasons to generate the intention to perform an action
As an example, the intention to move a wheel chair forward.

Transfer of the intention to the material representation of the human agent
By the upward transduction link *material representation of agent* to the component material representation of agent (to the property physical representation of move forward) and by the material link update material world to the component material world.

Execution of the material world
Causation of the EEG pattern from the property physical representation of move forward.

Observation of the EEG pattern by the computer agent
By the upward transduction link *symbolic effectuation of world* to the component symbolic representation of material world (to the property physical representation of move forward) and by the symbolic link observation results to the component computer agent.

Reasoning of the computer agent
Interpretation of the observed EEG pattern; deriving the action to be performed.

Transfer of the computer agent's action to the material world
By the symbolic link observations and actions to the component symbolic representation of material world and by the downward transduction link *material representation of world* to the component material world.

Execution of the action in the material world
The action move forward of the chair is executed, followed by causation of the movement of the human agent's body.

Observation of the new position by the human agent
By the upward transduction link *symbolic effectuation of world* to the component symbolic representation of material world and by the symbolic link observation results to the component human agent.

10. Discussion

Most paradigms for modelling or explanation assume that one scientific context or theory can be used for an explanation or model. For example, laws of chemistry should be used to explain bacterial behaviour. Or, cognitive theories should be used to explain behaviour of human or animal agents. In many cases phenomena occur that are too complex to be covered by one scientific context, also taken into account that theories within such contexts are often preliminary and still under development. An explanatory perspective sometimes advocated for such complex phenomena is explanatory pluralism. According to this perspective an explanation can consist of parts of a different signature, for example, a (partial) physical explanation and a (partial) mentalistic explanation. Each of these partial explanations is insufficient to explain the whole phenomenon, but together, if they are composed according to some form of interaction, they do explain the whole.

In this paper it was addressed how for such explanations the different types of interaction between mind and matter of an agent and the material world can be modelled in a conceptually and semantically sound manner, and how the overall explanation is composed from the parts, using these interactions. The hybrid agent architecture introduced can be used to model, explain and simulate a variety of phenomena in which multiple mind–matter interactions occur. It covers both a sub-model for the agent (simulating its mental processes) and a sub-model for the material world (simulating its physical processes). The semantic relations between the two sub-models are formalized as dual representation relations. In the model it is systematically taken into account that the agent's mind has a materialization in the form of a brain.

This hybrid agent modelling approach enables to model at the knowledge level integrated processes where physical- and cognitive-level descriptions of an agent influence each other. The usual approaches that either exploit cognitive style models or models for the physical dynamics do not cover such interactions. Validation of this hybrid model can be undertaken for a specific application. For example, for the application to *E.coli*'s behaviour, validation has taken place by means of experts from cell biochemistry; see Jonker *et al.* (2001, 2002). Each part of the knowledge within the model that (dynamically) relates beliefs, desires, intentions has been associated to chemical relationships that have been validated and confirmed. For application to humans such a validation may be more difficult, and maybe limited to validation of a coarser grain size in the sense that only rough knowledge is available that damage in a certain brain area as a whole relates to disturbances of certain cognitive functions. Other approaches to hybrid cognitive modelling combine symbolic modelling approaches (for the so-called explicit cognition) and connectionist approaches (for the so-called implicit cognition); for example see Sun (1994, 2000*b*); Sun and Alexandre (1997) and Sun and Bookman (1994). Although these types of approaches have a two-level perspective on cognition in common with our approach, in contrast to this work, our model is a formally specified knowledge-level model and does not commit to further implementation techniques.

Most parts of the specification of the model are generic; although the example instantiations that are used to illustrate the model are kept rather simple, the generic part of the model can be (re)used to simulate a variety of phenomena in which multiple mind–matter interactions occur. The component-based design method DESIRE supports that specific components in the model can be replaced by other components without affecting the rest of the model. For example, more sophisticated memory models can replace the rather simplistic model used as an illustration in this paper.

For further work, the approach presented in this paper may be of importance for the following.

(1) Foundational questions from a philosophical and logical perspective (cf. Bickle, 1998; Kim, 1998).
(2) Research in cognitive psychology, neuro-physiology, and their relation (cf. Bickle, 1998; Bechtel & Mundale, 1999).
(3) Research in biological context on modelling and explanation of intracellular processes and bacterial behaviour.

(4) Application to dynamic multi-agent domains in which agents can be created and killed (cf. Brazier *et al.*, 2001).
(5) Applications using direct brain–computer interfaces (cf. Birbaumer *et al.*, 1999, 2000; Levine *et al.*, 1999, 2000).

The relevance of the model for each of these three areas will be explained. The perspective of explanatory pluralism and its foundations is a central issue for which the approach presented in this paper is relevant (cf. McCauley, 1996; McCauley & Bechtel, 2001; Schouten & Looren de Jong, 2001; Looren de Jong, 2002). Moreover, an interesting more specific foundational philosophical and logical issue is the semantics of dual representation relations (see also Hofstadter, 1979). Both from a static and from a dynamic perspective further questions can be formulated and addressed—e.g. the further development of a foundation of semantic attachments and reflection principles (Weyhrauch, 1980) in the context of dual representation relations, and especially in dynamically changing mental and physical states (cf. Bickhard, 1993; Port & Gelder, 1995). Another question is the semantically sound integration of (qualitative and quantitative) simulation techniques and (temporal) logical modelling.

Cognitive and neuro-physiological models can be semantically integrated using the model introduced in this paper. The presented generic model can be instantiated by existing models of both kinds, and provides an integrative framework to glue partial explanations together to an explanation of a whole (cf. Bechtel & Mundale, 1999). A useful test for existing philosophical approaches to the mind–body problem (e.g. such as described by Bickle, 1998; Kim, 1998) is to investigate the possibility to operationalize them using the presented model.

Among the applications of the model are agents capable of planning and executing life affecting actions, such as giving birth and killing (other) agents; for an application in this area, see Brazier *et al.* (2001). These capabilities are essential for Internet agents that can decide on the fly to create new agents to assist them in their tasks and removing these agents after completion of the task they were created for.

The substantial progress made in the area of direct brain–computer interfacing (cf. Birbaumer *et al.*, 1999, 2000; Levine *et al.*, 1999, 2000) leads to high expectations for the development of applications of direct brain–computer interfacing in the area of Rehabilitation Engineering in the near future (e.g. Robinson, 2000). For modelling such applications involving direct brain–computer interfacing using the multiple mind-matter interaction model presented in this paper, two agents are involved: the human agent, and the supporting computer agent, for example related to a wheel chair. It was shown that the model can be used to clarify the different types of mind–matter interaction in this two-agent example in a transparent manner.

References

BARRINGER, H., FISHER, M., GABBAY, D., OWENS, R. & REYNOLDS, M. (1996). *The Imperative Future: Principles of Executable Temporal Logic*. New York: John Wiley & Sons, Research Studies Press Ltd.

BECHTEL, W. B. & MUNDALE, J. (1999). Multiple realisability revisited: linking cognitive and neural sciences. *Philosophy of Science*, **66,** 175–207.

BICKHARD, M.H. (1993). Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence*, **5**, 285–333.

BICKLE, J. (1998). *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.

BIRBAUMER, N., GHANAYIM, N., HINTERBERGER, T., IVERSEN, I., KOTCHOUBEY, B., KÜBLER, A., PERELMOUTER, J., TAUB, E., & FLOR, H. (1999). A spelling device for the paralysed. *Nature,* **398,** 297–298.

BIRBAUMER, N., KÜBLER, A., GHANAYIM, N., HINTERBERGER, T., PERELMOUTER, J., KAISER, J., IVERSEN, I., KOTCHOUBEY, B., NEUMANN, N., & FLOR, H. (2000). The thought translation device (TTD) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering*, **8,** 190–192.

BRAZIER, F. M. T., DUNIN-KEPLICZ, B., JENNINGS, N. R., & TREUR, J. (1995). Formal specification of multi-agent systems: a real-world case. In V. LESSER, Ed. *Proceedings of the 1st International Conference on Multi-agent Systems, ICMAS'95*, pp. 25–32. Cambridge, MA: MIT Press. Extended version in *International Journal of Cooperative Information Systems*, **6**(1997), 67–94 special issue on formal methods in cooperative information systems: multi-agent systems, M. HUHNS & M. SINGH, Eds).

BRAZIER, F. M. T., JONKER, C. M., & TREUR, J. (1998). Principles of compositional multi-agent system design. In J. CUENA Ed. *Proceedings of the 15th IFIP World Computer Congress, WCC'98, Conference on Information Technology and Knowledge Systems, IT&KNOWS'98*, pp. 347-360. Extended version in *Data and Knowledge Engineering*, **41**(2002), 1–28.

BRAZIER, F. M. T., JONKER, C. M., & TREUR, J. (2000). Compositional design and reuse of a generic agent model. *Applied Artificial Intelligence Journal*, **14,** 491–538. Shorter version in *Proceedings of the Knowledge Acquisition Workshop, KAW'99,* Banff, 1999.

BRAZIER, F. M. T., JONKER, C. M., TREUR, J., & WIJNGAARDS, N. J. E. (2001). Deliberative evolution in multi-agent systems. *Journal of Software Engineering and Knowledge Engineering*, **11,** 559–581. Preliminary version in J. GERO, Ed. *Artificial Intelligence in Design '00, Proceedings of the 6th International Conference on AI in Design, AID'2000*, pp. 663–650. Dordrecht: Kluwer Academic Publishers.

BRAZIER, F. M. T., TREUR, J., WIJNGAARDS, N. J. E., & WILLEMS, M. (1999). Temporal semantics of compositional task models and problem solving methods. *Data and Knowledge Engineering,* **29,** 17–42. Preliminary version in B.R. GAINES, & M.A. MUSEN, Eds. *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-based Systems Workshop, KAW'96*, pp. 15/1–15/17. Calgary: SRDG Publications, Department of Computer Science, University of Calgary.

CHANG, C. C., & KEISLER, H. J. (1973). *Model Theory*. Amsterdam: North-Holland.

DALEN, D. van (1980). *Logic and Structure*. Berlin: Springer-Verlag.

DENNETT, D. C. (1987). The International Stance, MIT Press, Cambridge, Mass.

DENNETT, D. C. (1991). Real Patterns. *The Journal of Philosophy*, **8,** 27–51.

ENGELFRIET, J., & TREUR, J. (1995). Temporal theories of reasoning. In C. MACNISH, D. PEARCE & L. M. PEREIRA Eds. *Logics in Artifical Intelligence, Proceedings of the 4th European Workshop on Logics in Artificial Intelligence, JELIA'94,* Berlin: Springer-Verlag. Also in *Journal of Applied Non-Classical Logics*, **5** (1995), 239–261.

HOFSTADTER, D. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books.

HODGES, W. (1993). *Model Theory*. Cambridge: Cambridge University Press.

JONKER, C.M., SNOEP, J. L., TREUR, J., WESTERHOFF, H. V., & WIJNGAARDS, W. C. A. (2001). Temporal modelling of the dynamics of intentional bacterial behaviour. In: R. COTTAM & R. PATON, Eds. *Proceedings of the 4th International Workshop on Information Processing in Cells and Tissues, IPCAT 2001,* pp. 177–188.

JONKER, C. M., SNOEP, J. L., TREUR, J., WESTERHOFF, H. V., & WIJNGAARDS, W. C. A. (2002). Putting intentions into cell biochemistry: an artificial intelligence perspective. *Journal of Theoretical Biology*, **214**, 105–134.

JONKER, C. M., & TREUR, J. (1997). Modelling an agent's mind and matter. In M. BOMAN, W. VAN DE VELDE Eds. *Multi-agent Rationality, Proceedings of the 8th European Workshop on*

Modelling Autonomous agents in a Multi- agent World, MAAMAW'97, *Lecture Notes in Artifical Intelligence* Vol. 1237, pp. 210-233. Berlin: Springer-Verlag.

Kahn, D., &Westerhoff, H. V.(1991). Control theory of regulatory cascades. *Journal of Theoretical Biology*, **153,** 255–885.

Kim, J. (1996). *Philosophy of Mind.* Boulder, Co: Westview Press.

Kim, J. (1998). *Mind in a Physical world: an Essay on the Mind–Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: an architecture for general intelligence. *Artificial Intelligence*, **33,** 1–64.

Levine, S. P., Huggins, J. E., Bement, S. L., Kushwaha, R. K., Schuh, L. A., Rohde, M. M., Passaro, E. A., Ross, D. A., Elisevich, K. V., & Smith B. J. (2000). A direct brain interface based on event-related potentials. *IEEE Transactions on Rehabilitation Engineering*, **8,** 180–185.

Levine, S. P., Huggins, J. E., Bement, S. L., Kushwaha, R. K., Schuh, L. A., Ross, D. A., & Rohde, M. M. (1999). A direct brain interface based on identification of electrocorticogram patterns related to voluntary actions. *Journal of Clinical Neurophysiology*, **16,** 439–447.

Lindsay, P. H., & Norman, D. A. (1977). *Human Information Processing*. New York: Academic Press.

Looren de Jong, H. (2002). *Levels of explanation in biological psychology*. Report, Department of Psychology, Vrije Universiteit Amsterdam.

McCauley, R. N. (1996). Explanatory pluralism and the coevolution of theories in science. In R.N. McCauley, Ed. *The Churchlands and Their Critics*, 17–47. Oxford: Blackwell.

McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and the heuristic identity theory. *Theory and Psychology*, **11**, 737–754.

Nagel, E. (1961). The Structure of Science. Harcourt, Brace & World, New York.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, **2**, 135–184.

Newell, A., Rosenbloom, P. S., & Laird, J. E. (1989). Symbolic architectures for cognition. In M.I. Posner, Ed. *Foundations Cognitive Science*, pp. 93–132. Cambridge, MA: MIT Press.

Port, R. F., & Gelder, T.van, Eds. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: MIT Press.

Posner, M. I., Ed. (1989). *Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1979). Do mental events have durations? *Behavioral and Brain Sciences*, **2,** 277–278.

Pylyshyn, Z. W. (1984). *Computation and Cognition: Towards a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1989). Computing in cognitive science. In M.I. Posner, Ed. *Foundations of Cognitive Science*, pp. 49–92. Cambridge, MA: MIT Press.

Rao, A. S., & Georgeff, M. P. (1991). Modelling rational agents within a BDI-architecture. In Allen, J., Fikes, R. & Sandewall, E. Eds. *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, KR'91*, pp. 473–484. Los Altos, CA: Morgan Kaufmann.

Robinson, C. J. (2000). A commentary on brain–computer interfacing and its impact on rehabilitation science and clinical applicability (introduction to special section on brain–computer interfaces). *IEEE Transactions on Rehabilitation Engineering*, **8,** 161–162.

Rohwer, J. M., Schuster, S., Westerhoff, H. V. (1996). How to recognize monofunctional units in a metabolic system. *Journal of Theoretical Biology*, **179,** 213–228.

Schilling, C. H., Letscher, D., & Palsson, B. Ø. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, **203,** 229–248.

Schouten, M. K. D., & Looren de Jong, H. (2001). Explanatory pluralism and heuristic identification: some explorations in behavioral genetics. *Theory and Psychology,* **11,** 799–810.

Simon, H. A., & Kaplan, C. A. (1989). *Foundations of Cognitive Science*, In M.I., Posner, Ed. pp. 1–48. Cambridge, MA: MIT Press.

Smith, B. C. (1982). *Reflection and semantics in a procedural language*. MIT Comp. Science Tech. Report 272, Cambridge, MA.

SMITH, B. C. (1985). Prologue to reflection and semantics in a procedural language. In R.J. BRACHMAN & H.J. LEVESQUE Eds. *Readings in Knowledge Representation*, pp. 31-40. Los Altos, CA: Morgan Kaufman.

SUN, R. (1994). *Integrating Rules and Connectionism for Robust Commonsense Reasoning*. New York, NY: John Wiley and Sons.

SUN, R. (2000a). Symbol grounding: a new look at an old idea. *Philosophical Psychology*, **13,** 149–172.

SUN, R. (2000b). *Duality of the Mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

SUN, R., & F. ALEXANDRE Eds. (1997). *Connectionist-Symbolic Integration*. Mahwah, NJ: Lawrence Erlbaum Associates.

SUN, R., & BOOKMAN, L. Eds. (1994). *Computational Architectures Integrating Neural and Symbolic Processes.* Dordrecht: Kluwer Academic Publishers.

TARSKI, A., Der Wahrheitsbegriff in den formalisierten Sprachen. *StudiaPhilosophica*, **1,** 261–405.(English translation: A. TARSKI, (1956). *Logic, Semantics, Metamathematics.* Oxford: Oxford University Press).

TREUR, J. (1991). Declarative functionality descriptions of interactive reasoning modules. In H. BOLEY & M.M. RICHTER Eds. *Processing Declarative Knowledge, Proceedings of the International Workshop PDK-91, Lecture Notes in Artifical Intelligence*, Vol. 567, pp. 221–236. Berlin: Springer-Verlag. Extended version to appear in *International Journal of Intelligent Systems*, **17**(2002), 645–686.

TREUR, J. (1994). Temporal semantics of meta-level architectures for dynamic control of reasoning. In L. FRIBOURG & F. TURINI Eds. *Logic Program Synthesis and Transformation—Meta-programming in Logic, Proceedings of the 4th International Workshop on Meta-programming in Logic, META'94, Lecture Notes in Computer Science,* Vol. 883, pp. 353–376. Berlin: Springer Verlag. Extended version in *International Journal of Intelligent Systems,* **17**(2002) 545–568.

WEYHRAUCH, R. W. (1980). Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence,* **13**, 133–170.

Paper accepted for publication by associate editor M. Musen.