On the Use of Reduction Relations to Relate Different Types of Agent Models

Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence, Agent Systems Research Group
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
treur@cs.vu.nl http://www.cs.vu.nl/~treur

Abstract. This paper focuses on relationships between agent models and their physical realisations. Approaches on reduction from philosophical literature are analysed in a formalised manner, and extended by incorporating context-dependency w.r.t. specific makeups for their realisations. It is shown how these context-dependent reduction approaches can be translated into each other and how they can be applied to relate agent models.

Keywords. Reduction relation, agent model, context-dependent, bridge law, interpretation mapping, functional reduction

1. Introduction

Agent models can be designed at different levels of abstraction. For example, the well-known BDI-model (e.g., [34]) makes use of higher-level cognitive concepts such as beliefs, desires and intentions. Agent models can also be defined on the basis of the world's dynamics, and described by concepts at lower levels, for example, physical, chemical, or neurological concepts; e.g., [9, 10]. In order to ground models for embodied agents in a physical, chemical or neurological context, often the focus is on their interaction as a coupled system with the environment; e.g., [2, 11, 12, 22, 22]. However, they can be related to physical reality in a still more fundamental manner when the model of their internal functioning is fully immersed in a model of the world's dynamics, and to this end concepts from a lower level are used in the model, or it is indicated how the concepts used in the model relate to such lower-level concepts. In this way cognition can be addressed by an artifical life like approach; e.g., [10, 22, 32, 36]. This allows to model other types of mind-matter interaction, such as an agent that takes drugs which change its internal functioning (e.g., antidepressiva that affect neuro-transmitter levels), electrostimulation therapies, or brain-computer interfacing (e.g., [22]).

It is an interesting challenge to explore how exactly a higher-level agent model can be immersed in a lower-level model of the world's dynamics. A related fundamental question is whether this can be done in exactly one way, or in multiple ways. For example, is cognitive functioning depending on a neurological realisation, or is a different realisation also possible? Within the philosophical literature reduction approaches are proposed to relate higher-level and lower-level descriptions (theories). In this paper three of these reduction approaches are analysed on applicability: the bridge law approach [31], the functional approach [14, 24, 25], and the interpretation mapping approach [37].

The paper is organised as follows. After a brief introduction to to reduction in Section 2, Section 3 shows how the three reduction approaches can be refined to incorporate the notion of a specific makeup, thus obtaining context-dependent variants that allow multiple realisation. Here a context is defined as a specific makeup within a lower-level theory. In Section 4 in a further comparative analysis it is shown under which conditions and how the approaches can be related to each other by mutual translations. In Section 5 a practical case study illustrates the applicability of the different approaches to relate a higher-level agent models to lower-level models.

2. Reduction Approaches

Work on reduction can be found in a wide variety of publications in the philosophical literature; see, for example [3, 4, 5, 14, 24, 25, 31]. Reduction addresses relationships between descriptions of two different levels, usually indicated by a higher-level theory T_2 (e.g., a cognitive theory) and a lower-level or base theory T_1 (e.g., a neurological theory). A specific reduction approach provides a particular *reduction relation*: a way in which each higher-level property a (an expression in T_2) can be related to a lower-level property b (an expression in T_1), this b is often called a *realiser* for a. Reduction approaches differ in how these relations are defined. In the classical *bridge law reduction* approach, following [31] reduction relations are specified by (biconditional) bridge principles $a \leftrightarrow b$ that relate the expressions a in the language of a higher-level theory T_2 in a one-to-one manner to expressions b in the language of the lower-level theory T_1. As an alternative Kim [25], pp.

98-102 describes *functional reduction* based on functionalisation of a target property a in T_2 in terms of its causal task C (specifying its causal relationships to other properties) and relating it to a state property b in T_1 performing this causal task C. For functional reduction the reduction relations are not required to be one-to-one, thus allowing multiple realisation. A third notion to define reduction relations is a *(relative) interpretation mappings* (e.g., [3], [17], pp. 201-263; [22], pp. 61-65; [37]). These approaches relate the two theories T_2 and T_1 based on a mapping φ from expressions a of T_2 to expressions b of T_1, in the sense that $b = \varphi(a)$.

Below in Sections 2.1 to 2.3 these three reduction approaches are introduced in more detail. In Section 2.4 they are compared.

2.1 Bridge law reduction

Nagel's classical definition of a reduction relation between a theory T_2 and a theory T_1 is based on a set of specified (biconditional) bridge principles as follows:

a) A *bridge principle* (or *bridge law*) is a principle or law connecting an expression of T_2 to an expression of T_1. A bridge principle is *biconditional* if it has the form of a logical equivalence $a \leftrightarrow b$ where a is an expression of T_2 and b an expression of T_1.

b) A theory T_2 is *bridge law reducible* to T_1 on the basis of a specified set of bridge principles $a_i \leftrightarrow b_i$ connecting each basic expression a_i of T_2 with a unique expression b_i of T_1 if and only if all laws $L(a_1, \ldots, a_k)$ of T_2 are derivable from the laws of T_1 augmented with the bridge principles $a_i \leftrightarrow b_i$ for the a_i that occur in the law $L(a_1, \ldots, a_k)$, i.e., within the language of T_1 augmented with symbols for the basic statements $a_1, \ldots, a_k$ it holds:

$$T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\} \vdash L(a_1, \ldots, a_k)$$

Here $T \vdash A$ denotes that A is derivable from T. Note that the notation $L(a_1, \ldots, a_k)$ is used here to indicate how a more complex statement is built as a proposition from subformulae $a_1, \ldots, a_k$. Furthermore, note that in logical terms the theory $T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\}$ is a *definitional extension* of T_1 obtained by adding (new) symbols[1] for a_i to T_1 and their definitions given by the biconditional bridge principles; see, for example [22], pp. 57-61; [17], p. 60. Every definitional extension is a *conservative extension*; therefore for all statements α in the language of T_1 it holds

$$T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\} \vdash \alpha \text{ iff } T_1 \vdash \alpha$$

See, for example [22], pp. 41, 57-61; [17], pp. 59-60, 66. From this it can be derived that the criterion for bridge law reduction has an equivalent formulation:

$$\begin{aligned} & T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\} \vdash L(a_1, \ldots, a_k) \\ \Leftrightarrow\ & T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\} \vdash L(b_1, \ldots, b_k) \\ \Leftrightarrow\ & T_1 \vdash L(b_1, \ldots, b_k) \end{aligned}$$

Here the last equivalence follows since $L(b_1, \ldots, b_k)$ belongs to the language of T_1, and the theory

$$T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\}$$

is a conservative extension of T_1. So, summarising, the following are equivalent formulations for the criterion of bridge law reduction for a law $L(a_1, \ldots, a_k)$ of T_2 where

$$a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k$$

are bridge principles:

(i) $T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow T_1 \cup \{a_1 \leftrightarrow b_1, \ldots, a_k \leftrightarrow b_k\} \vdash L(a_1, \ldots, a_k)$

(ii) $T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow T_1 \vdash L(b_1, \ldots, b_k)$

The key concept here is the specification of the bridge principles, which relate each basic expression of T_2 in a unique manner to an expression of T_1. In practice, these bridge principles have to be biconditional to have the possibility of deriving nontrivial T_2-laws from T_1 laws, thereby satisfying b). For example, suppose $F \rightarrow G$ is derivable from T_2, and $F^* \rightarrow G^*$ is a derivable from T_1. Then by bridge principles $F \leftrightarrow F^*$ and $G \leftrightarrow G^*$ (since they are biconditional), the T_2-law $F \rightarrow G$ can be derived from the T_1-law $F^* \rightarrow G^*$. As an example, suppose for theory T_2 mental state properties a_1, a_2, observation *obs* and action *act* are given such that T_2 specifies that

$$\begin{aligned} & obs \rightarrow a_1 \\ & a_1 \rightarrow a_2 \\ & a_2 \rightarrow act \end{aligned}$$

holds. Assume that b_1, b_2 (e.g., indicating activation states of sensory and prepatory neurons) are realisers of a_1, a_2 in T_1, using bridge principles expressed by

$$\begin{aligned} & a_1 \leftrightarrow b_1 \\ & a_2 \leftrightarrow b_2 \end{aligned}$$

whereas $b_1 \rightarrow b_2$ is derivable within T_1. Then $a_1 \rightarrow a_2$ is derivable from $b_1 \rightarrow b_2$ (which is derivable from T_1) and the given bridge laws.

2.2 Functional reduction

Bridge law reduction has difficulties to handle cases where multiple realisations are possible. To cover such multiple realisation cases, the notion of *functional reduction* was developed (e.g., [25], pp. 98-102; [24], pp. 19-23, 97-103), described in brief as follows by Kim [25], pp. 101-102:

[1] Note that symbol names in different theories are assumed distinct.

STEP 1 [FUNCTIONALIZATION OF THE TARGET PROPERTY *a*]
Property *a* to be reduced is given a functional definition of the following form:
Having $a =_{def.}$ having some property or other *P* (in the reduction base domain) such that *P* performs causal task *C*.
For a functionally defined property *a*, any property in the base domain that fits the causal specific-ation definitive of *a* (that is, a property that performs causal task *C*) is called a 'realizer' of *a*.

STEP 2 [IDENTIFICATION OF THE REALIZERS OF *a*]
Find the properties (or mechanisms) in the reduction base that perform the causal task *C*.

STEP 3 [DEVELOPING AN EXPLANATION THEORY]
Construct a theory that explains how the realizers of *a* perform task *C*.

Note that the process of functionalisation specifies a (higher-level) property as a *second-order property*: a property about other (lower-level, base) properties. Instead of the term 'causal task', sometimes also terms such as 'causal role' or 'functional role' are used. The functionalization of mental state properties makes them relational: they are specified by how they (causally) relate to other properties. Kim ([14], pp. 200-202) uses a similar idea to solve problems in the area of representational content of mental state properties. Kim [25] claims that the specification *C* of the functional role of *a* is well-suited for a reduction relation, for *a* to be mapped onto lower level properties and their causal relationships within the lower-level theory. A subtle issue is that the specification of functional role *C* often involves other mental state properties, which have their own causal tasks. In [23], p. 105, an example is discussed. In such a case a joint functionalisation of a set of related mental state properties can be achieved based on the Ramsey-Lewis method, following [27], [33]; see also [23], pp. 105-107. This method works as follows. Continuing the earlier example, suppose again mental state properties a_1, a_2, observation *obs* and action *act* are given such that

$$obs \rightarrow a_1$$
$$a_1 \rightarrow a_2$$
$$a_2 \rightarrow act$$

holds. Then a joint causal task specification for a_1 and a_2 can be expressed as:

$$C(M_1, M_2) =_{def.} (obs \rightarrow M_1) \,\&\, (M_1 \rightarrow M_2) \,\&\, (M_2 \rightarrow act)$$

Here the (second-order) variable M_1 is used to indicate a_1's causal role and M_2 to indicate a_2's causal role. A physicalist functionalist assumes that the properties that the joint causal role specification *C* takes to exist are physical properties; that is, the variables $M_1, M_2, \ldots$ range over physical state properties, and are often indicated by $P_1, P_2, \ldots$. Using existential quantification for these variables over the domain of state properties of T_1, the following functional definition of having mental state property a_1, resp, a_2 (conceptualised as a second-order property) is obtained:

$$\text{Having } a_1 =_{def} \exists P_1,P_2 \,[\, C(P_1, P_2) \,\&\, P_1 \text{ holds }]$$
$$\text{Having } a_2 =_{def} \exists P_1,P_2 \,[\, C(P_1, P_2) \,\&\, P_2 \text{ holds }]$$

The fact that b_1, b_2 are considered to be realisers of a_1, a_2 in T_1 based on functional reduction is expressed by

$$C(b_1, b_2)$$

whereas $b_1 \rightarrow b_2$ is derivable within T_1.

Given a mental state property *a*, to define the role of *a* by the Ramsey-Lewis method, a joint causal role description *C* has to be chosen in such a way that: (1) *a* occurs in *C,* and (2) for every mental state property occurring in *C*, its functional role description occurs in *C*. In other words: starting with *a*, its functional role description has to be added, for each new mental state property occurring in the resulting specification, also its functional role description has to be added, and so on and so forth. This follows the transitive closure of the relation 'occurs in the functional role description of' between mental state properties.

An assumption is that a joint causal role specification $C(P_1, \ldots, P_k)$ can be identified such that it covers the relevant state properties $a_1, \ldots, a_k$ of theory T_2, and at least one instantiation of it within T_1 exists: $\exists P_1, \ldots, P_k$ $T_1 \vdash C(P_1, \ldots, P_k)$. This joint causal role specification is kept fixed. Any law $L(a_1, \ldots, a_k)$ derivable from T_2 relates to all T_1-expressions $L(P_1, \ldots, P_k)$ for $P_1, \ldots, P_k$ in T_1 for which $C(P_1, \ldots, P_k)$ is derivable in T_1. In more precise terms, all of these T_1-expressions $L(P_1, \ldots, P_k)$ for $P_1, \ldots, P_k$ for which $C(P_1, \ldots, P_k)$ is derivable from T_1, are themselves derivable from theory T_1:

$$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow$$
$$\forall P_1, \ldots, P_k \,[\, T_1 \vdash C(P_1, \ldots, P_k) \Rightarrow T_1 \vdash L(P_1, \ldots, P_k) \,]$$

Note again that (1) here the arguments indicated by $P_1, \ldots, P_k$ refer to state properties within T_1; they are second-order variables. Furthermore, note that (2) the collection of them written down need not be the exact set of the state properties occurring both in *C* and in *L*, as these sets may differ for *C* and *L*. The $P_1, \ldots, P_k$ are chosen in such a way that the intersection of state properties occurring in *C* and *L* is included. These notes are relevant throughout the paper.

2.3 Reduction based on an interpretation mapping

The basic idea of an interpretation of a theory T_2 in a theory T_1 is that expressions *a* from T_2 are related to expressions *b* from T_1 by an appropriate mapping φ. This mapping from the expressions of T_2 to expressions of T_1 is specified in such a manner that if an expression or law *L* can be derived from T_2, then $\varphi(L)$ can be derived from T_1:

$T_2 \vdash L \Rightarrow T_1 \vdash \varphi(L)$

Again continuing the earlier example, suppose mental state properties a_1, a_2, observation *obs* and action *act* are given such that

$obs \rightarrow a_1 \rightarrow a_2 \rightarrow act$

holds. The fact that b_1, b_2 are considered to be realisers of a_1, a_2 in T_1 based on an interpretation mapping is expressed by

$\varphi(a_1) = b_1$
$\varphi(a_2) = b_2$

whereas $b_1 \rightarrow b_2$ is derivable within T_1. Additional conditions on the mapping φ often express that it is defined in an effective manner and is compositional, i.e., that it preserves the compositional logical structure of the expression:

$\varphi(a_1 \wedge a_2) = \varphi(a_1) \wedge \varphi(a_2)$
$\varphi(a_1 \vee a_2) = \varphi(a_1) \vee \varphi(a_2)$
$\varphi(a_1 \rightarrow a_2) = \varphi(a_1) \rightarrow \varphi(a_2)$
$\varphi(\neg a) = \neg \varphi(a)$
$\varphi(\forall x\, A) = \forall x\, \varphi(A)$
$\varphi(\exists x\, A) = \exists x\, \varphi(A)$

These rules for preservation of compositional structure can be used to define an interpretation mapping for more complex formulae in an inductive manner, taking the mapping of atoms as a point of departure. Note that T_2-atoms may be mapped onto T_2-atoms, but can equally well be mapped onto more complex T_2-formulae. Sometimes also atoms are mapped according to their structure. For example, the mapping of an atom *R(f(x), g(y))* with relation symbol *R*, function symbols *f* and *g* and constants *v* and *w* can be based on mappings between symbols $R \rightarrow R'$, $f \rightarrow f'$, $g \rightarrow g'$, $v \rightarrow v'$, $w \rightarrow w'$ to obtain

$\varphi(R(f(v), g(w))) = R'(f'(v'), g'(w'))$

In such a way an interpretation mapping can be based on an *ontology mapping*, i.e., a mapping of the different basic ontological elements used. A variation on this is when the predicate *R* is mapped onto a more complex formula $\alpha(x, y)$, and

$\varphi(R(f(v), g(w))) = \alpha(f'(v'), g'(w'))$

For more details on interpretations in formal logical literature, see, for example [37]; [22], pp. 61-65; [17], pp. 201-263. For more discussion on variants of the interpretation mapping approach within philosophical literature, see for example [3].

2.4 Comparison of the three reduction approaches

The three classical reduction approaches considered above differ in the way in which they specify reduction relations to relate expressions *a* in T_2 and *b* in T_1, as shown in Table 1: based on bridge principles, on causal role specifications, and on interpretation mappings, respectively.

Table 1
Three approaches to reduction: overview

	relations between state properties		*relations between laws*	
bridge law reduction	biconditional bridge law relating *a* and *b*	$a \leftrightarrow b$	law *L* of T_2 is deriv-able from laws of T_1 plus bridge principles	$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow$ $T_1 \cup \{a_i \leftrightarrow b_i\} \vdash L(a_1, \ldots, a_k)$
functional reduction	joint causal role specification *C* for $a_1, \ldots, a_k$ is satis-fied by $b_1, \ldots, b_k$	$C(b_1, \ldots, b_k)$ holds with $a_i \equiv \exists P_1, .. P_k$ $C(P_1, .., P_k) \& P_i$	law *L* of T_2 relates to a collection of T_1-expressions all deriv-able from laws of T_1	$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow \forall P_1, \ldots, P_k$ $[\, T_1 \vdash C(P_1, \ldots, P_k) \Rightarrow$ $T_1 \vdash L(P_1, \ldots, P_k)]$
interpretation mapping	mapping φ relating *a* and *b*	$b = \varphi(a)$	law *L* of T_2 maps on T_1-expression deriv-able from laws of T_1	$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow$ $T_1 \vdash \varphi(L(a_1, \ldots, a_k))$

An interesting question is in how far they are still equivalent in some sense: unlike the different formats to specify reduction relations, is it possible to translate them into each other? For the approaches as described in the literature, for the general case the answer on this question is negative. A main reason for this is that the approaches treat multiple realisation differently. For bridge law reduction the biconditionality criterion implies uniqueness of the realiser up to equivalence: suppose for *a* two bridge principles $a \leftrightarrow b$ and $a \leftrightarrow b'$ are given with non-equivalent *b* and *b'*, then by symmetry and transitivity of the logical biconditional relation $\leftrightarrow$ within theory

$T_1 \cup \{a \leftrightarrow b, a \leftrightarrow b'\}$

it holds $b \leftrightarrow b'$, which would contradict that $b \leftrightarrow b'$ does not hold within T_1. This means that bridge law reduction implies unique realisation; it shows why multiple realisation is not covered adequately by bridge law reduction. This is different for functional reduction. In

that case, by the quantification over $P_1, \ldots, P_k$ multiple realisation is covered, but in an implicit manner, i.e., without explicitly specifying the different options for realisation. For reduction based on interpretation mappings the situation is still different. If an interpretation mapping is specified, this indicates just one realisation, so in that sense this does not support multiple realisation. However, it does not exclude the existence of different (non-equivalent) realisers, as bridge laws do due to their biconditional character. It is quite well possible that two interpretation mappings φ and φ' exist such that not for all a it holds $\varphi(a) \leftrightarrow \varphi'(a)$. This shows that reduction based on interpretation mappings does not exclude multiple realisation, but using just one mapping does not provide a specification covering it. However, if multiple interpretation mappings are used, it will be able to cover multiple realisation in an explicit manner. Such a type of extension will be made in the next section. It will be set up in a more general form by making context-dependency of a set of realisers explicit so that it can be applied to the other two approaches as well.

In summary, from the three approaches mentioned, functional reduction is able to handle multiple realisation, but in an implicit manner. The interpretation mapping approach can be extended when multiple mappings are taken into account, so that multiple realisation is covered in an explicit manner. Bridge law reduction is not able to handle multiple realisation. However, as a way out, in [14], pp. 233-236, Kim briefly sketches how what he calls a *local* or *structure-restricted* form of bridge law reduction, can handle multiple realisation. His suggestion is to relativise bridge principles $a \leftrightarrow b$ to

$$S \rightarrow (a \leftrightarrow b)$$

by adding an extra parameter S indicating the context of a specific system or makeup of an organism. Below in Section 3, in line with [38], it is shown how a variation on this idea of context-dependent reduction can be worked out in more detail for each of the three reduction approaches considered. Thus refined variants are obtained making multiple realisation explicit by reference to the context-dependency of a specific realisation. It will turn out that systematic relationships between these three refined approaches exist (see Section 4 for mutual translations).

3. Context-Dependent Reduction

In context-dependent reduction as introduced in [38], the aim is to identify a set of contexts and to relate the different realisations to these contexts. When contexts are defined in a sufficiently fine-grained manner, within one context the realisation can be unique. In this case, from an abstract viewpoint contexts can be seen as a form of parameterisation of the different possible realisations. For example, in Cognitive Science such a grouping could be based on species, i.e., groups of organisms with (more or less) the same makeup. If mental state properties (for example, having a certain sensory representation) are assumed that can be shared between, for example, biological organisms and robot-like systems it may be useful to allow contexts that are described within different base theories. Therefore in the context-dependent reduction approach developed, a collection of (base) theories $\mathcal{T}_1$ is assumed and for each theory T in $\mathcal{T}_1$ a set of contexts $\mathcal{C}_T$, such that each particular context[2] of an organism or system is formally described by a pair (T, S) of a specific theory T in $\mathcal{T}_1$ together with a specific context S in $\mathcal{C}_T$. The contexts S are assumed to be descriptions in the language of T and consistent with T. The theories T in $\mathcal{T}_1$ and contexts S in $\mathcal{C}_T$ can be used to distinguish the different realisations that are possible. Below it is shown how this can be done for the three approaches considered. Note that when the collection of theories $\mathcal{T}_1$ is taken a singleton $\{T_1\}$ consisting of one theory T_1 and the set of contexts $\mathcal{C}_{T_1}$ is taken a singleton $\{S\}$ consisting of the empty specification $S = \phi$, then the original general reduction approach is obtained.

3.1 Context-dependent bridge law reduction

For the bridge law reduction approach, the set of realisers that exists within one context S for a theory T in $\mathcal{T}_1$, is expressed by context-dependent biconditional bridge laws parameterised by a theory T in $\mathcal{T}_1$ and a context S in $\mathcal{C}_T$, specified by

$$a_1 \leftrightarrow b_{1,T,S}, \ldots, a_k \leftrightarrow b_{k,T,S}$$

Given such a parameterised specification, the context-dependent criterion of bridge law reduction for a law $L(a_1, \ldots, a_k)$ derivable from T_2 can be formulated (in two equivalent manners) by[3]:

(i) $T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow$
$\forall T \in \mathcal{T}_1 \; \forall S \in \mathcal{C}_T$
$T \cup S \cup \{a_1 \leftrightarrow b_{1,T,S}, \ldots, a_k \leftrightarrow b_{k,T,S}\} \vdash L(a_1, \ldots, a_k)$

(ii) $T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow$
$\forall T \in \mathcal{T}_1 \; \forall S \in \mathcal{C}_T \; T \cup S \vdash L(b_{1,T,S}, \ldots, b_{k,T,S})$

Note that context-dependent bridge law reduction implies unique realisers (up to equivalence) per context: from $a \leftrightarrow b_{T,S}$ and $a \leftrightarrow b'_{T,S}$ it follows that $b_{T,S}$ and $b'_{T,S}$ cannot be non-equivalent in $T \cup S$. So to obtain context-dependent bridge law reduction in cases of multiple

[2] For the sake of shortness a context (T, S) is often indicated just by S.

[3] Note that the notation $L(a_1, \ldots, a_k)$ is used to indicate how a more complex statement is built as a proposition from subformulae $a_1, \ldots, a_k$. Furthermore, the theory $T \cup S \cup \{a_1 \leftrightarrow b_{1,T,S}, \ldots, a_k \leftrightarrow b_{k,T,S}\}$ is a *definitional extension* of $T \cup S$ obtained by adding (new) symbols for a_i to T; see, for example [9], p. 60; [35], p. 57-61. Every definitional extension is a *conservative extension*; therefore for all statements α in the language of T it holds $T \cup S \cup \{a_1 \leftrightarrow b_{1,T,S}, \ldots, a_k \leftrightarrow b_{k,T,S}\} \vdash \alpha$ if and only if $T \cup S \vdash \alpha$; see, e.g., [9], pp. 59-60, 66; [35], p. 41, 57-61.

realisation, the contexts are defined with a grain-size such that per context a unique realisation exists.

3.2 Context-dependent functional reduction

For a given collection of context theories $\mathcal{T}_1$ and sets of contexts $\mathcal{C}_T$, for context-dependent functional reduction a first criterion is that a joint causal role specification[4] $C(P_1, ..., P_k)$ can be identified such that it covers all relevant state properties of theory T_2, and for each theory T in $\mathcal{T}_1$ and context S in $\mathcal{C}_T$ at least one instantiation of it within T exists:

$$\forall T \in \mathcal{T}_1 \, \forall S \in \mathcal{C}_T \, \exists P_1, ..., P_k \;\; T \cup S \vdash C(P_1, ..., P_k).$$

The second criterion for context-dependent functional reduction, concerning laws is

$$T_2 \vdash L(a_1, ..., a_k) \;\Rightarrow\; \forall T \in \mathcal{T}_1 \, \forall S \in \mathcal{C}_T \, \forall P_1, ..., P_k \\ [\, T \cup S \vdash C(P_1, ..., P_k) \;\Rightarrow\; T \cup S \vdash L(P_1, ..., P_k) \,]$$

In general this notion of context-dependent functional reduction may still allow multiple realisation within one theory and context. However, by choosing contexts with an appropriate grain-size it can be achieved that within one given theory and context unique realisation occurs. The *unique realisation context criterion* (also called *strictness criterion*) expresses this as follows. For each T in $\mathcal{T}_1$ and context S in $\mathcal{C}_T$ there exists a unique set of instantiations realising the joint causal role specification $C(P_1, ..., P_k)$, or formally:

$$\forall T \in \mathcal{T}_1 \, \forall S \in \mathcal{C}_T \;\; \exists P_1, ..., P_k \\ [\, T \cup S \vdash C(P_1, ..., P_k) \;\&\; \\ \forall Q_1, ..., Q_k \;\; [\; T \cup S \vdash C(Q_1, ..., Q_k) \Rightarrow \\ T \cup S \vdash P_1 \leftrightarrow Q_1 \;\&\; ... \;\&\; P_k \leftrightarrow Q_k \,] \,]$$

This guarantees per theory T and context S unique realisers, parameterised by T and S. When also this third criterion is satisfied, a form of reduction is obtained that we call *strict* context-dependent functional reduction. When the strictness criterion is satisfied, the universally quantified form for relations between laws is equivalent to an existentially quantified variant:

$$T_2 \vdash L(a_1, ..., a_k) \;\Rightarrow\; \forall T \in \mathcal{T}_1 \, \forall S \in \mathcal{C}_T \; \exists P_1, ..., P_k \\ [T \cup S \vdash C(P_1, ..., P_k) \;\&\; T \cup S \vdash L(P_1, ..., P_k) \,]$$

3.3 Context-dependent interpretation

To obtain a form of context-dependent interpretation, the notion of interpretation mapping is generalised to a multi-mapping, parameterised by contexts. A *context-dependent interpretation* of a theory T_2 in a collection of theories $\mathcal{T}_1$ with sets of contexts $\mathcal{C}_T$ specifies for each theory T in $\mathcal{T}_1$ and context S in $\mathcal{C}_T$ an appropriate mapping $\varphi_{T,S}$ from the expressions of T_2 to expressions of T: a multi-mapping

$$\varphi_{T,S} \, (T \in \mathcal{T}_1, S \in \mathcal{C}_T)$$

from theory T_2 to theories T in $\mathcal{T}_1$ parameterised by theories T in $\mathcal{T}_1$ and contexts S in $\mathcal{C}_T$. Such a multi-mapping is a context-dependent interpretation mapping when it satisfies the property that if a law L can be derived from T_2, then for each T in $\mathcal{T}_1$ and context S in $\mathcal{C}_T$ the statement $\varphi_{T,S}(L)$ can be derived from $T \cup S$:

$$T_2 \vdash L \;\Rightarrow\; \forall T \in \mathcal{T}_1 \, \forall S \in \mathcal{C}_T \;\; T \cup S \vdash \varphi_{T,S}(L)$$

Usually the mappings are assumed compositional with respect to logical connectives, as expressed in Table 2.

Table 2 Compositionality of an interpretation mapping for logical connectives

$\varphi_{T,S}(A_1 \wedge A_2) = \varphi_{T,S}(A_1) \wedge \varphi_{T,S}(A_2$
$\varphi_{T,S}(A_1 \vee A_2) = \varphi_{T,S}(A_1) \vee \varphi_{T,S}(A_2)$
$\varphi_{T,S}(A_1 \rightarrow A_2) = \varphi_{T,S}(A_1) \rightarrow \varphi_{T,S}(A_2)$
$\varphi_{T,S}(\neg A) = \neg \varphi_{T,S}(A)$
$\varphi_{T,S}(\forall x \, A(x)) = \forall x \, \varphi_{T,S}(A(x))$
$\varphi_{T,S}(\exists x \, A(x)) = \exists x \, \varphi_{T,S}(A(x))$

Note that also here within one theory T in $\mathcal{T}_1$ and context S in $\mathcal{C}_T$ multiple realisation is possible, expressed as the existence of two essentially different interpreation mappings $\varphi_{T,S}$ and $\varphi'_{T,S}$, i.e. such that in $T \cup S$ it may not hold that $\varphi_{T,S}(a) \leftrightarrow \varphi'_{T,S}(a)$. However, an additional *strictness criterion* to obtain unique realisation per context is formulated as follows: when for any given theory T in $\mathcal{T}_1$ and context S in $\mathcal{C}_T$ two interpretation mapping $\varphi_{T,S}$ and $\varphi'_{T,S}$ are possible, then for all a it holds that

$$T \cup S \vdash \varphi_{T,S}(a) \leftrightarrow \varphi'_{T,S}(a)$$

When this additional criterion is satisfied as well, the interpretation is called a *strict* context-dependent interpretation.

[4] This specifies the causal relationships of these properties to each other and to other (external) properties; this can be obtained by the Ramsey-Lewis method as described in [23], [27], [33].

4. Mutual Translations

In this section it is shown how the context-dependent interpretation mapping approach can be related to the other two context-dependent approaches by mutual translations.

4.1 Relating bridge law reduction interpretation

In this subsection it is shown how bridge law reduction can be translated into reduction based on a strict interpretation mapping and vice versa.

4.1.1 From interpretation to bridge law reduction

Suppose a strict interpretation mapping $\varphi_{T,S}$ is given, which is assumed compositional. For each basic expression a of T_2 specify the bridge principle

$$a_i \leftrightarrow b_{i,T,S} \quad \text{with} \quad b_{i,T,S} = \varphi_{T,S}(a_i)$$

If $L(a_1, \ldots, a_k)$ is law derivable from T_2 involving state properties $a_1, \ldots, a_k$, then

$$T \cup S \vdash \varphi_{T,S}(L(a_1, \ldots, a_k)).$$

By compositionality of φ it follows that

$$T \cup S \vdash L(\varphi_{T,S}(a_1), \ldots, \varphi_{T,S}(a_k)).$$

Therefore it follows

$$T \cup S \vdash L(b_{1,T,S}, \ldots, b_{k,T,S}).$$

This shows that the criterion (ii) for bridge law reduction is fulfilled. Note that it is needed to assume the interpretation to be strict. If the same translation would be done for two essentially different non-strict interpretations $\varphi_{T,S}$ and $\varphi'_{T,S}$, it would lead to contradictions.

4.1.2 From bridge law reduction to interpretation

For a translation the other way around, assume bridge principles

$$a_i \leftrightarrow b_{i,T,S}$$

are given for the basic expressions a_i of T_2, such that the bridge law reduction criterion (ii) is fulfilled:

$$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow T \cup S \vdash L(b_{1,T,S}, \ldots, b_{k,T,S})$$

Define the mapping $\varphi_{T,S}$ as follows. For each basic expression a_i of T_2, based on the given bridge principle $a_i \leftrightarrow b_{i,T,S}$, define

$$\varphi_{T,S}(a_i) = b_{i,T,S}$$

For more complex expressions extend this by compositionality as described in Table 2. For this mapping $\varphi_{T,S}$, from $T_2 \vdash L(a_1, \ldots, a_k)$ by the bridge law reduction criterion (ii) it follows by compositionality:

$$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow$$
$$T \cup S \vdash L(\varphi_{T,S}(a_1), \ldots, \varphi_{T,S}(a_k)) \Rightarrow$$
$$T \cup S \vdash \varphi_{T,S}(L(a_1, \ldots, a_k)).$$

Therefore the criterion for an interpretation mapping is fulfilled.

Note that the two translations from bridge law reduction to interpretation and from interpretation to bridge law reduction as given are each others' inverse. Moreover, note that the context-dependent interpretation obtained from bridge law reduction is strict. When within a given theory and context an essentially different interpretation would be possible, this could be translated into bridge laws as well, which would lead to a contradiction.

4.2 Relating interpretation to functional reduction

Next it is shown how functional reduction can be translated into an interpretation and vice versa.

4.2.1 From functional reduction to interpretation

Let a theory T in $\mathbf{\mathcal{T}}_1$ and a context S in $\mathbf{C}_T$ be given. Take a joint causal role specification $C(P_1, \ldots, P_k)$ for the basic state properties $a_1, \ldots, a_k$ of T_2. Suppose $L(a_1, \ldots, a_k)$ is derivable from T_2 is given and the functional reduction criterion holds:

$$T_2 \vdash L(a_1, \ldots, a_k) \Rightarrow \forall P_1, \ldots, P_k$$
$$[T \cup S \vdash C(P_1, \ldots, P_k) \Rightarrow T \cup S \vdash L(P_1, \ldots, P_k)]$$

Pick an arbitrary set $b_1, \ldots, b_k$ of realisers satisfying $C(P_1, \ldots, P_k)$ in $T \cup S$, and define

$$\varphi_{T,S}(a_i) = b_i$$

For more complex expressions extend this by compositionality as described in Table 2. For $L(a_1, \ldots, a_k)$ derivable from T_2, by the functional reduction criterion it holds:

$$T \cup S \vdash C(\varphi_{T,S}(a_1), \ldots, \varphi_{T,S}(a_k)) \Rightarrow$$
$$T \cup S \vdash L(\varphi_{T,S}(a_1), \ldots, \varphi_{T,S}(a_k))$$

As the antecedent holds due to the choice of the mapping, it follows that

$$T \cup S \vdash L(\varphi_{T,S}(a_1), \ldots, \varphi_{T,S}(a_k)).$$

By the compositional definition of $\varphi_{T,S}$ for more complex expressions, as before, it follows:

$$T \cup S \vdash \varphi_{T,S}(L(a_1, \ldots, a_k)).$$

Therefore the interpretation mapping criterion is fulfilled for the chosen mapping $\varphi_{T,S}$. Note that a mapping $\varphi_{T,S}$ as

defined above fully depends on the chosen set of realisers of the joint causal role specification $C(P_1, ..., P_k)$. This may result in a collection of possible mappings for each instantiation of $P_1, ..., P_k$ satisfying $C(P_1, ..., P_k)$ in $T \cup S$. This can be avoided by assuming the additional criterion of unique realisation context:

$$\exists P_1, ..., P_k \ [T \cup S \vdash C(P_1, ..., P_k) \ \& \ \forall Q_1, ..., Q_k \ [\ T \cup S \vdash C(Q_1, ..., Q_k) \ \Rightarrow \ T \cup S \vdash P_1 \leftrightarrow Q_1 \ \& \ ... \ \& \ P_k \leftrightarrow Q_k \ ] \ ]$$

Using this (i.e., assuming strict context-dependent functional reduction), per theory T and context S a unique interpretation mapping is found. This is a strict interpretation mapping, because any essentially different interpretation mapping would provide another set of realisers within the same context.

4.2.2 From interpretation to functional reduction

Suppose a context-dependent interpretation mapping $\varphi_{T,S}$ ($T \in \mathbf{\mathcal{T}}_1$, $S \in \mathbf{C}_T$) is given, which is assumed compositional. Moreover, let $L(a_1, ..., a_k)$ be derivable from T_2 . Let a theory T in $\mathbf{\mathcal{T}}_1$ and a context S in $\mathbf{C}_T$ be given. Then by the interpretation mapping criterion it holds $T \cup S \vdash \varphi_{T,S}(L(a_1, ..., a_k))$ and hence by the compositionality assumption it holds

$$T \cup S \vdash L(\varphi_{T,S}(a_1), ..., \varphi_{T,S}(a_k)).$$

Assume that a joint causal role specification is given by $C(a_1, ..., a_k)$, which means $T_2 \vdash C(a_1, ..., a_k)$ holds. The interpretation mapping criterion applied to $C(a_1, ..., a_k)$ provides

$$T_2 \vdash C(a_1, ..., a_k) \ \Rightarrow T \cup S \vdash \varphi_{T,S}(C(a_1, ..., a_k)).$$

By the compositionality assumption it holds

$$\varphi_{T,S}(C(a_1, ..., a_k)) = C(\varphi_{T,S}(a_1), ..., \varphi_{T,S}(a_k)).$$

Hence from $T_2 \vdash C(a_1, ..., a_k)$ it follows

$$T \cup S \vdash C(\varphi_{T,S}(a_1), ..., \varphi_{T,S}(a_k)).$$

Therefore, if the variables $P_1, ..., P_k$ are instantiated by $\varphi_{T,S}(a_1), ..., \varphi_{T,S}(a_k)$, it holds

$$T \cup S \vdash C(P_1, ..., P_k) \ \& \ T \cup S \vdash L(P_1, ..., P_k)$$

Now this only shows that for some instantiation of the variables $P_1, ..., P_k$ the functional reduction criterion holds, not for all instantiations. In fact for the general case the following weaker existential criterion is implied:

$$T_2 \vdash L(a_1, ..., a_k) \Rightarrow \exists P_1, ..., P_k \ [T \cup S \vdash C(P_1, ..., P_k) \ \& \ T \cup S \vdash L(P_1, ..., P_k) \ ]$$

This weaker existential criterion is (only) equivalent to the stronger universal criterion when it is assumed that exactly one unique set of instantiations of the variables $P_1, ..., P_k$ is possible within context S for theory T such that $C(P_1, ..., P_k)$ holds (unique realisation criterion). Therefore to obtain a faithful translation it is assumed that $\varphi_{T,S}$ is a strict context-dependent interpretation. In such a case if an essentially different realising instantiation of $C(P_1, ..., P_k)$ would be possible, this would lead to an essentially different interpretation mapping (see previous translation), which is excluded by the criterion of strictness. Thus by the translation discussed above, both (equivalent) universal and existential versions of the criterion for strict context-dependent functional reduction are satisfied:

$$T_2 \vdash L(a_1, ..., a_k) \Rightarrow \forall T \in \mathbf{\mathcal{T}}_1, S \in \mathbf{C}_T \ \forall P_1, ..., P_k \ [T \cup S \vdash C(P_1, ..., P_k) \Rightarrow T \cup S \vdash L(P_1, ..., P_k) \ ]$$

$$T_2 \vdash L(a_1, ..., a_k) \Rightarrow \forall T \in \mathbf{\mathcal{T}}_1, S \in \mathbf{C}_T \ \exists P_1, ..., P_k \ [T \cup S \vdash C(P_1, ..., P_k) \ \& \ T \cup S \vdash L(P_1, ..., P_k) \ ]$$

5. Relating Agent Models

In this section the concepts discussed above are illustrated for a case study involving a higher-level cognitive model *CM* and two lower-level models: a neurological model *NM* and a biochemical model *BM*. The neurological model *NM* will be described by a general neurological theory *NT* and a specific makeup *NS*, describing a specific context. Similarly, the biochemical model *BM* is described by a general biochemical theory *CT* and a specific makeup *CS*, which describes (in a simplified form) the context of the bacterium *E. coli*. Based on the neurological theory *NT* and biochemical theory *CM* and the neural example context *NS* and the biochemical context *BS* of *E. coli*, the context-dependent interpretation ($\varphi_{NT,NS}$, $\varphi_{BT,BS}$) for the cognitive model *CM* is defined:

$$T_2 = CM, \ \mathbf{\mathcal{T}}_1 = \{NT, CT\}, \mathbf{C}_{NT} = \{NS\}, \mathbf{C}_{CT} = \{CS\}.$$

5.1 The cognitive model CM

This model plays the role of the higher-level theory. It describes a simple cognitive process which depending on observations on two world facts *s1* and *s2* makes a choice between two actions *a1* and *a2*. It is specified as follows:

worldfact(X) $\rightarrow$ *observed(X)*
observed(X) $\rightarrow$ *belief(X)*
belief(s1) & not *belief(s2)* $\rightarrow$ *intention(a1)*
belief(s2) $\rightarrow$ *intention(a2)*
intention(a1) & *belief(s1)* $\rightarrow$ *performed(a1)*
intention(a2) & *belief(s2)* $\rightarrow$ *performed(a2)*
performed(a1) & *worldfact(s1)* $\rightarrow$ *worldfact(e1)*
performed(a2) & *worldfact(s2)* $\rightarrow$ *worldfact(e2)*

5.2 The neurological model NM

For the neurological model a situation is taken involving two objects. When a cube is seen and no sphere, it will be taken, when a sphere is seen and no cube, it will be taken. When both are seen, only the sphere is taken. The neurological model NM used consists of the general laws specified in (simplified) neurological theory *NT* and a specific neural makeup described by *NS*.

Neurological theory *NT* Activations of neurons propagate via connections through synapses with positive (excitory) or negative (inhibitory) effects. In case of multiple connections to one neuron, the effect is combined, and activation takes place when this combined input is above the neuron's threshold. When a sensor stimulus occurs that is connected to a neuron, then this neuron is activated, when the input is above its threshold. When the combined input for an action is above its threshold, then this action occurs. This is formalised as:

connectedto(X, Y, pos) & *activated(X)*
& *threshold(Y, v)* & $v < 1 \rightarrow$ *activated(Y)*
connectedto(X1, X2, Y, pos) & *activated(X1)* & *activated(X2)*
& *threshold(Y, v)* & $v < 2 \rightarrow$ *activated(Y)*
connectedto (X1, Y, pos) & *connectedto(X2, Y, neg)*
& *activated(X1)* & not *activated(X2)*
& *threshold(Y, v)* & $v < 1 \rightarrow$ *activated(Y)*
occurs(X) $\rightarrow$ *seeing(X)*
activated(take(X)) $\rightarrow$ *having(X)*

Note that here the predicate *connectedto* is used to represent all combined positive input for a neuron and all combined negative input for a neuron. Moreover, note that for convenience in the last line some world relations have been included.

Neural makeup *NS* (see Fig. 1):

connectedto(seeing(cube), SN1, pos)
connectedto(seeing(sphere), SN2, pos)
connectedto(SN2, MN2, take(sphere), pos)
connectedto(SN1, MN1, take(cube), pos)
connectedto(SN1, MN1, pos)
connectedto(SN2, MN2, pos)
connectedto(SN2 MN1, neg)
threshold(SN1, 0.5)
threshold(SN2, 0.5)
threshold(MN1, 0.5)
threshold(MN2, 0.5)
threshold(take(cube), 1.5)
threshold(take(sphere), 1.5)

5.3 Mapping the cognitive model onto the neural model

Given the cognitive model *CM* and the neurological model *NM*, the next step is to relate them by a reduction relation. As in Section 4 it was shown how the different context-dependent reduction approaches c*an be translated into each other, it is only shown for one of them: the interpretation mapping approach. The* cognitive model *CM* is mapped onto the neurological model *NM* by the interpretation mapping $\varphi_{NT,NS}$ defined by (and extended to more complex propositions in a compositional manner according to Table 2):

$\varphi_{NT,NS}$(*observed(s1)*) = *seeimg(cube)*
$\varphi_{NT,NS}$(*observed(s2)*) = *seeing(sphere)*
$\varphi_{NT,NS}$(*belief(s1)*) = *activated(SN1)*
$\varphi_{NT,NS}$(*belief(s2)* = *activated(SN2)*
$\varphi_{NT,NS}$(*intention(a1)*) = *activated(MN1)*
$\varphi_{NT,NS}$(*intention(a2)*) = *activated(MN2)*
$\varphi_{NT,NS}$(*performed(a1)*) = *activated(take(cube))*
$\varphi_{NT,NS}$(*performed(a2)*) = *activated (take(sphere))*

For example, the relation

belief(s1) & not *belief(s2)* $\rightarrow$ *intention(a1)*

of the cognitive model is mapped by $\varphi_{NT,NS}$ as follows:

$\varphi_{NT,NS}$(*belief(s1)* & not *belief(s2)* $\rightarrow$ *intention(a1)*)
= $\varphi_{NT,NS}$(*belief(s1)* & not *belief(s2)*) $\rightarrow$ $\varphi_{NT,NS}$(*intention(a1)*)
= $\varphi_{NT,NS}$(*belief(s1)*) & $\varphi_{NT,NS}$(not *belief(s2)*) $\rightarrow$ $\varphi_{NT,NS}$(*intention(a1)*)
= $\varphi_{NT,NS}$(*belief(s1)*) & not $\varphi_{NT,NS}$(*belief(s2)*) $\rightarrow$ $\varphi_{NT,NS}$(*intention(a1)*)
= *activated(SN1)* & not *activated(SN2)* $\rightarrow$ *activated(MN1)*

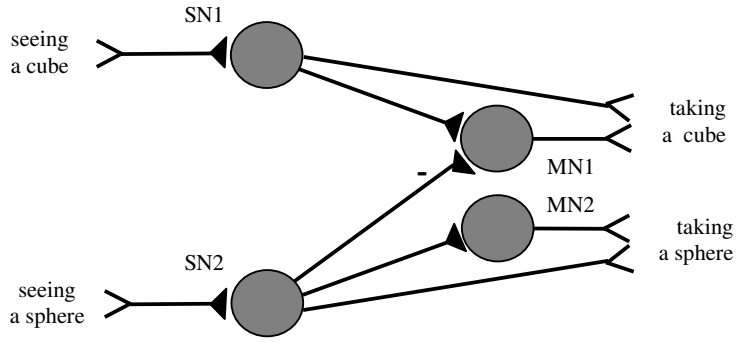

Fig. 1. Neural makeup *NS*

From $NT \cup NS$ the following relationships can be derived:

occurs(cube) → *seeing(cube)*
occurs(sphere) → *seeing(sphere)*
seeingcube → *activated(SN1)*
seeingsphere → *activated(SN2)*
activated(SN2) → *activated(MN2)*
activated(SN1) & not activated(SN2) → *activated(MN1)*
activated(MN1) & activated(SN1 → *activated(take(cube))*
activated(MN2) & activated(SN1) → *activated(take(sphere))*
activated(take(cube)) & occurs(cube) → *having(cube)*
activated(take(sphere) & occurs(sphere) → *having(sphere)*

These relationships are exactly the mapped relationships from *CM*, formally: $\varphi_{NT,NS}(CM)$. This shows that the criterion for interpretation is satisfied.

5.4 The biochemical model BM

Within cell biology causal chains are known in the form of chemical pathways from the environment to within the cell. For example, such causal chains justify to interpret the presence of an internal concentration of CRPcAMP above a certain level as an indicator for 'glucose being absent in the external environment', and of the internal presence of a certain concentration of lactose as an indicator for 'lactose being present in the external environment'. This shows ways in which a cell is able to build and maintain internal states that can be interpreted as a world model, or its beliefs about the world. Intentions can be considered to be present in the cell in that, depending on the observed environment it is able to make an informed choice (for preparation of an action) between alternatives of specific import action to provide resources for a specific type of metabolism. See [19, 20] for more detailed models for intracellular processes underlying bacterial behaviour. For the simplified example considered here, the biochemical theory *BT* consists of general biochemical laws indicating how certain types of substances in general can react with each other. The makeup *BS* of *E.coli* specifies the presence of a specific cell membrane, a water-like fluid inside of appropriate temperature, and the presence of specific substances within the cell, such as DNA.

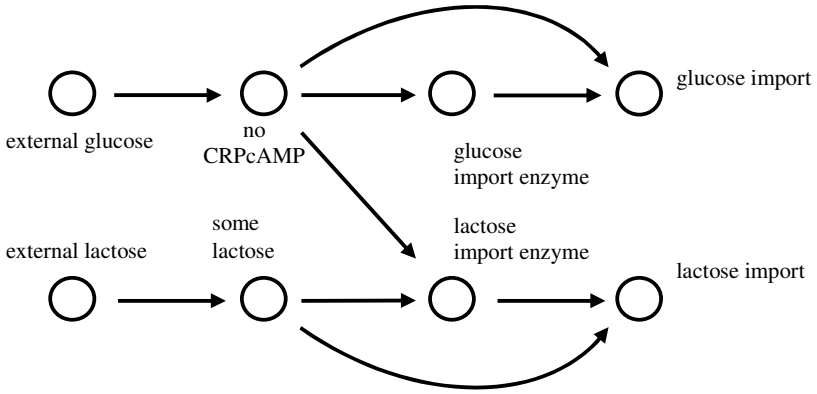

Fig. 2. Derived biochemical relationships for *E. coli*

5.5 Mapping the cognitive onto the biochemical model

The cognitive model *CM* can be mapped onto the biochemical model *BM* by the interpretation mapping $\varphi_{BT,BS}$ defined by:

$\varphi_{NT,NS}$(*observation(s1)*) = *external lactose*
$\varphi_{NT,NS}$(*observation(s2)*) = *external glucose*
$\varphi_{BT,BS}$(*belief(s1)* = *some lactose*
$\varphi_{BT,BS}$(*belief(s2)* = *no CRPcAMP*
$\varphi_{BT,BS}$(*intention(a1)*) = *lactose import enzyme*
$\varphi_{BT,BS}$(*intention(a2)*) = *glucose import enzyme*
$\varphi_{NT,NS}$(*action(a1)*) = *import lactose*
$\varphi_{NT,NS}$(*action(a2)*) = *import glucose*

Again the mapping is extended for more complex propositions in a compositional manner (see Table 2). For example, the relation

belief(s1) & not *belief(s2)* → *intention(a1)*

of the cognitive model is mapped by $\varphi_{BT,BS}$ as follows:

$\varphi_{BT,BS}$(*belief(s1)* & not *belief(s2)* → *intention(a1)*)
= $\varphi_{BT,BS}$(*belief(s1)* & not *belief(s2)*) → $\varphi_{BT,BS}$(*intention(a1)*)
= $\varphi_{BT,BS}$(*belief(s1)*) & $\varphi_{BT,BS}$(not *belief(s2)*) → $\varphi_{BT,BS}$(*intention(a1)*)
= $\varphi_{BT,BS}$(*belief(s1)*) & not $\varphi_{BT,BS}$(*belief(s2)*) → $\varphi_{BT,BS}$(*intention(a1)*)
= *some lactose* & not *no CRPcAMP* → *lactose import enzyme*

From $BT \cup BS$ the following relationships can be derived. For example, for sensory processes the following derivable relationships describe that the external presence of glucose and or lactose leads to the presence of the related internal indicators.

external glucose → *no CRPcAMP*
external lactose → *some lactose*
no CRPcAMP → *glucose import enzyme*
glucose import enzyme → *glucose import*
lactose import enzyme → *lactose import*
some lactose & not *no CRPcAMP* → *lactose import enzyme*

The relationships for action generation derivable from $BT \cup BS$ cover transcription (DNA affecting the presence of mRNA), translation (mRNA affecting the presence of enzyme), cathalysis (enzymes affecting the related import reactions), and the effects of (co)factors in these steps (also see Fig. 2). These relationships are the mapped relationships from *CM*, formally: $\varphi_{BT,BS}(CM)$. This illustrates the fulfilment of the criterion for interpretation.

6. Discussion

Agents described by a higher-level model can have different physical realisations. In this paper it was shown how some of the approaches on reduction available in philosophical literature can be applied to relate such a higher-level agent model to its physical realisations. Below a number of aspects and implications of this work are discussed.

6.1 Context-dependency of physical realisation

As the three approaches to reduction do not treat multiple realisation in an explicit manner, refined variants of all three approaches were used making multiple realisation explicit by reference to the context-dependency of a specific realisation. The notion of context-dependency distinguishes the general rules or laws of an underlying theory from more specific aspects such as a particular makeup of an agent, for example, the general rules for neural systems in contrast to a particular neural architecture. It turned out to be possible to obtain systematic relationships between the three refined context-dependent reduction approaches, in the form of mutual translations between them. The treatment as presented abstracts from the dynamic aspects (following what is usually done in the philosophical literature mentioned). A topic for future work is to make these dynamic aspects explicit, by considering a form of temporal logic.

In a case study it was shown how based on the machinery developed a cognitive agent model can be related both to a realisation by a neural model and by a biochemical model. Here the neural and the biochemical model each consist of a generic part specifying general laws or rules, and a specific part specifying the particular makeup considered. The higher-level cognitive model unifies basic properties of the two different lower-level models and describes them in a more abstract manner.

6.2 Evaluation in other case studies

Two other case studies can be mentioned that illustrate the approach presented here. The first one addresses the relationship between a biological and a cognitive agent model for criminal behaviour; see [7] for details. Here the notion of interpretation mapping is used to obtain clarification of the cognitive model in relation to underlying biological factors. As an example of a relation between basic concepts, a desire for aggressive behaviour is related to the biological factor testosterone level. Other relationships can be found in Table 3.

Table 3 Relationships for the criminal behaviour case study

Cognitive Conceptualisation	Biological Conceptualisation
sensitivity_for_stimuli(v)	chemical_state(serotonin, v)
preparedness_to_act(v)	chemical_state(adrenalin, v)
preparedness_to_safety(v)	chemical_state(oxytocine, v)
desire_for_ strong_stimuli(v)	brain_state_for_stimulation(v')
desire_for_aggressiveness(v)	chemical_state(testosterone, v)
desire_to_act(v)	chemical_state(adrenalin, v)
desire_to_act_safely(v)	chemical_state(oxytocine, v)
desire_for_impulsiveness(v)	chemical_state(bloodsugar, v)

Another case study addresses the relationship between cognitive and biological agent models for emotion reading; for details, see [28]. For example the cognitive state srs(s) for sensory representation of a stimulus s is related to activation of sensory neurons SRN(s) for s. For some more relationships, see Table 4.

Table 4 Relationships for the emotion reading case study

Cognitive Conceptualisation	Biological Conceptualisation
srs(s)	activated(SRN(s), 1)
preparation_state(f, v)	activated(PN(f), v)
emotion(e, v)	activated(EN(e), v)
effector_state(f, v)	effector_state(f, v)
sensor_state(f, v)	sensor_state(f, v)
imputation(s, e)	$\exists v$ $v \geq 0.75$ & activated(RN(s, e), v)

6.3 Mind-matter interaction

Mind-matter interaction plays an important role in a number of applications; for example, see [22]. One application area of such mind-matter relationships concerns the use of drugs which affect cognitive functioning, as often plays an important role in the functioning of humans, for example persons suffering from depression, psychiatric patients, or criminals with deviant brain structures; e.g., [6]. On the one hand dynamical system models exist that estimate the concentration of drugs in the blood (e.g., [18]), and on the

other hand models that describe cognitive functioning, but those models are rarely formally integrated or related to each other; e.g., [8]. Another application area for mind-matter relationships is brain-computer interfacing; see, for example, [14], [29], [30]. Here by monitoring the physical states of the brain, estimations are made of a human's cognitive states, and used for example, to control a machine or a wheel chair. For both application areas mentioned, to provide integrated formal models for the mind-matter interaction involved is an interesting challenge still to be addressed. By having (in addition) a realisation of a higher-level agent model, it becomes possible to incorporate mutual effects between the physical world and an agent's internal functioning. Within the framework presented here, effects of the world on the agent's functioning in principle can be modelled as a change from the agent's makeup S to a changed makeup S' (for example, a drug that affects the activation threshold of neurons, or inactivation of certain genes at the cell's DNA). In general such a changed makeup S' can be a realisation of the same or of another higher-level agent model.

6.4 Model-driven development of agent applications

Model-driven software development is an important recent development within software engineering; e.g., [15],[26], which also induces developments in application areas in the context of the Internet and software agents; e.g., [13], [16], [1]. For this type of application areas in particular reusable agent model libraries are being set up. Traditionally agent models have a symbolic, logical character, suited for qualitative applications. However, it is more and more recognized that quantitative applications based on numerical models are important to be addressed for agents as well, especially for applications dealing with continuous world dynamics and more complex adaptive types of behaviour; e.g., [9]. Therefore among agent models in libraries also such numerical models are to be included. Various examples of such numerical models can be found in the areas of neural (and complex adaptive) systems modelling. The scope of applications for agent systems can be substantially extended when such models are available in libraries in formats that enables reusability and integration. Reusability will be supported more when the formats used in such a library allow different views on the same model according to different levels of abstraction. A formalised context-dependent reduction relation (in the form of an interpretation mapping or bridge laws, or function reduction relation as described in the current paper) can be used to relate such views to each other. An example of an application area where different views and their relations may be relevant are embodied agents such as found in robotics, where both a physical model and a cognitive model may be useful descriptive means. Other examples which can benefit from such agent models concern application areas where adaptivity is important, such as adaptive and personalised Web applications.

6.5 Possibilities for automated support

The concepts and formalisations presented here can be used for model conceptualisation and specification, thereby relating ontologies used and defining different views taking such relationships into account. However, the conceptual machinery may also be useful by establishing relationships between existing agent models. The question can be put forward in how far automated support can be developed to verify whether a reduction relation between existing models holds; for example, to check whether a given (ontology) mapping provides an interpretation mapping, i.e., to check whether the condition

$$T_2 \vdash L \Rightarrow T \cup S \vdash \varphi_{T,S}(L)$$

is fulfilled for a given $T \in \boldsymbol{\mathcal{T}_1}$ and $S \in \boldsymbol{C_T}$. Here the properties L can be just taken as those defining the specification of T_2, which is a finite (and maybe not too large) set. For each of such properties L, to automatically check whether

$$T \cup S \vdash \varphi_{T,S}(L)$$

holds using available tools for model checking and/or theorem proving seems a feasible route to be explored. This is left for future work.

Acknowledgements

The paper has benefit from constructive comments by the anonymous reviewers.

References

[1] Agüero, J., Rebollo, M., Carrascosa, C., and Julián, V. (2009). Agent Design Using Model Driven Development. In: Demazeau, Y., Pavón, J., Corchado, J.M., Bajo, J. (eds.), Proc. of the 7th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS'09). Advances in Intelligent and Soft Computing, vol. 55. Springer Verlag, 2009, pp. 60-69.

[2] Bickhard, M.H. (1993). Representational Content in Humans and Machines. *J. of Experimental and Theoretical AI*, vol. 5, 1993, pp. 285-333.

[3] Bickle, J. (1992). Mental Anatomy and the New Mind-Brain Reductionism. *Philosophy of Science*, vol. 59, pp. 217-230.

[4] Bickle, J. (1998). *Psychoneural Reduction: The New Wave*. MIT Press, Cambridge, Mass.

[5] Bickle, J. (2003). *Philosophy and Neuroscience*. Kluwer Academic Publishers.

[6] Blair, R.J.R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition,* 14 (2005) 698–718.

[7] Bosse, T., Gerritsen, C., and Treur, J., Grounding a Cognitive Modelling Approach for Criminal Behaviour. In: S. Vosniadou, D. Kayser, A. Protopapas (eds.),

Proceedings of the Second European Cognitive Science Conference, EuroCogSci'07. Lawrence Erlbaum Associates, 2007, pp. 776-781.

[8] Bosse, T., Gerritsen, C.G., and Treur, J., Towards Integration of Biological, Psychological, and Social Aspects in Agent-Based Simulation of Violent Offenders. *Simulation Journal*. In press, 2009.

[9] Bosse, T., Jonker, C.M., and Treur, J., (2007). Simulation and Analysis of Adaptive Agents: an Integrative Modelling Approach. *Advances in Complex Systems*, vol. 10, 2007, pp. 335 - 357.

[10] Bosse, T., and Treur, J., Formalising Agency-Inducing Patterns in World Dynamics. In: Rocha, L.M., et al.. (eds.), *Artificial Life X: Proc. of the 10th International Conference*. MIT Press, 2006, pp. 546-552.

[11] Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge University Press.

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

[13] Claus, P. (2008). Semantic model-driven development of web service architectures. International Journal of Web Engineering and Technology, vol. 4, 2008, pp. 386-404.

[14] Dornhege, G., Millán, J. del, Hinterberger, T., McFarland, D., and Müller, K.R. (eds.), (2007). *Toward Brain-Computer Interfacing*. MIT Press, Cambridge, MA, 2007.

[15] Greenfield, J., and Short, K. (2005). *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*. Wiley Publishing.

[16] Grønmo, R., Skogan, D., Solheim, I., Oldevik, J. (2004). Model-Driven Web Services Development, International Journal of Web Services Research, vol. 1, 2004, pp. 1 - 13

[17] Hodges, W. (1993). *Model theory*. Cambridge University Press.

[18] Hoogendoorn, M., Klein, M., Memon, Z., and Treur, J., Formal Analysis of Intelligent Agents for Model-Based Medicine Usage Managment. In: Azevedo, L. and Londral, A.R. (eds.), *Proc. of the First International Conference on Health Informatics, HEALTHINF 2008*. INSTICC Press, 2008, pp. 148 - 155.

[19] Jonker, C.M., Snoep, J.L., Treur, J., Westerhoff, H.V., and Wijngaards, W.C.A. (2002). Putting Intentions into Cell Biochemistry: An Artificial Intelligence Perspective. *Journal of Theoretical Biology,* vol. 214, pp. 105-134.

[20] Jonker, C.M., Snoep, J.L., Treur, J., Westerhoff, H.V., Wijngaards, W.C.A., (2008). BDI-Modelling of Complex Intracellular Dynamics. *Journal of Theoretical Biology*, vol. 251, 2008, pp. 1–23.

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

[22] Jonker, C.M., and Treur, J., Modelling Multiple Mind-Matter Interaction. *International Journal of Human-Computer Studies*, vol. 57, 2002, pp. 165-214.

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

[24] Kim, J. (1998). *Mind in a Physical world*. MIT Press.

[25] Kim, J. (2005). *Physicalism, or Something Near Enough.* Princeton University Press, Princeton.

[26] Kleppe, A. (2003). MDA Explained, The Model-Driven Architecture: Practice and Promise. Addison-Wesley.

[27] Lewis, D.K. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, vol. 50, pp 249-258.

[28] Memon, Z.A., and Treur, J., Cognitive and Biological Agent Models for Emotion Reading. In: Jain, L., Gini, M., Faltings, B.B., Terano, T., Zhang, C., Cercone, N., Cao, L. (eds.), *Proceedings of the 8th IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT'08*. IEEE Computer Society Press, 2008, pp. 308-313.

[29] Millán, J. del R., Ferrez, P.W., Galán, F., Lew, E., and Chavarriaga, R. (2008). Non-Invasive Brain-Machine Interaction. *International Journal of Pattern Recognition and Artificial Intelligence*, 22:959–972.

[30] Millán, J. del R. (2003). Adaptive Brain Interfaces. *Communications of the ACM*, vol. 46, pp. 74–80

[31] Nagel, E. (1961). *The Structure of Science*, London: Routledge and Kegan Paul.

[32] Port, R.F., Gelder, T. van (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition.* MIT Press, Cambridge, Mass, 1995.

[33] Ramsey, F.P. (1929). *Theories*. In: Ramsey, F.P. (1931) *The Foundations of Mathematics and Other Essays* (R. B. Braithwaite, ed.), Routledge and Kegan Paul.

[34] Rao, A.S. & Georgeff, M.P. (1991). Modelling Rational Agents within a BDI-architecture. In: Allen, J., et al. (eds.), *Proc. of the 2nd Intern. Conf. on Principles of Knowledge Representation and Reasoning (KR'91),* Morgan Kaufmann, pp. 473-484.

[35] Schoenfield, J.R. (1967). *Mathematical Logic*. Addison-Wesley.

[36] Steels, L. & Brooks, R. (1995). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Erlbaum.

[37] Tarski, A., Mostowski, A., and Robinson, R.M. (1953). *Undecidable Theories*. North-Holland.

[38] Treur, J., Laws and Makeups in Context-Dependent Reduction Relations. In: Love, B.C., McRae, K., and Sloutsky, V.M. (eds.), *Proc. of the 30th Annual Conference of the Cognitive Science Society, CogSci'08*. Cognitive Science Society, Austin, TX, 2008, pp. 1752-1757.