Title
Developing Higher-Level Cognitive Theories by Reduction

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 28(28)

ISSN

1069-7977

Authors

Bosse, Tibor
Jonker, Catholijn M.
Treur, Jan

Publication Date

2006

Peer reviewed

Developing Higher-Level Cognitive Theories by Reduction

Tibor Bosse (tbosse@cs.vu.nl), Catholijn M. Jonker[1] (C.Jonker@nici.ru.nl), Jan Treur[2] (treur@cs.vu.nl)
Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

Abstract

Within philosophical literature, higher-level cognitive concepts such as free will, authorship of actions, and conscious control are often questioned. Neurological and biochemical mechanisms underlying human behavior provide alternative explanations of action. Reduction of cognitive states to neurophysiologic states shows that higher-level cognitive concepts in principle can be eliminated, replacing them by neurophysiologic concepts. In contrast, in this paper it is shown how reduction relations can be used in a constructive manner to strengthen the scientific foundation of higher-level cognitive concepts and further develop higher-level theories in which these concepts play a role.

Introduction

Reduction is an important theme within literature in the area of Cognitive Science, Philosophy of Science, and Philosophy of Mind; e.g., Kim (1996, 1998, 2005), Bickle (1998, 2003). One of the main perspectives advocated is physicalism: the idea that all processes (among which mental and biological processes), in one way or the other have a physical basis. Within Biology the strong development of biochemistry supports this perspective. For Cognitive Science, the strong development of neuroscience and its underlying biochemistry plays a similar role. These developments sometimes lead to a position that considers higher-level concepts, such as intention, free will, and consciousness, just illusions, and not usable in a scientific context; e.g., Wegner (2002).

A main question addressed in this paper is how to defend such human-like, higher-level concepts. One strategy is to criticise the existence of reduction relations; if these do not exist, it is impossible to relate higher-level concepts to lower-level ones, and in this way to eliminate the higher-level concepts; e.g., Bennett and Hacker (2003, pp. 355-366). This defensive strategy is not applicable in those cases where reduction relations have been or are being shown to exist. Another strategy is to put doubt on the quality of the higher-level theory, for example, by claiming that there do not exist cognitive laws that could be related to neurological laws; e.g., Bennett and Hacker (2003, p. 362). This strategy is not applicable in cases that the higher-level theory is assumed to be still under development. In this paper a different strategy is explored. This strategy takes the existence of reduction relations in the present or in the future as an assumption, and shows in some detail how they can be used to strengthen the usefulness and further development of the higher-level theory.

This position acknowledges the achievements of neuroscience and shows how to exploit these results where possible, in favour of higher-level concepts instead of against them. This paper shows how this perspective provides a scientific and philosophical foundation for higher-level notions. A practical method, based on formal techniques and tools, is proposed to support the perspective. The method will result, among others, in relational specifications of the functional role and of the representational content of these concepts; cf. Kim (1996, 1998, 2005). The method allows to check whether these specifications are logically coherent mutually, and with global behavioral properties, and, in as far as available from neuroscience, consistent with neurological theories. The resulting relational specifications can be said to specify meaning and underlying mechanisms for the higher-level concepts, and as such provide grounding of them both within the higher-level theory and within physical reality. Furthermore, the paper addresses different notions of explanation in terms of the higher-level theory and the lower-level theory. From a practical point of view, in a specific case study the predictive value of higher-level explanations is validated.

On the Use of Reduction

In the philosophical literature on reduction of scientific theories, in many cases the advantages for scientific practice of having a reduction relation between two theories are not addressed explicitly. In such cases, sometimes it is implicitly assumed that these advantages are based on the idea to use the lower-level theory instead of the higher-level the theory. For example, Kim (1996, pp. 214-216) emphasizes three advantages of reduction based on (biconditional) bridge principles: ontological simplification (following Smart, 1959), having to deal with fewer assumptions about the world, and providing explanations of the laws of the higher-level theory in terms of the lower-level theory. For example, if F and G are higher-level entities and F^* and G^* lower-level ones with biconditional bridge laws $F \leftrightarrow F^*$ and $G \leftrightarrow G^*$ (cf. Nagel, 1961), then F^* and G^* can be used instead of F and G.[3] An eliminative perspective on the use of reduction provokes resistance from those who defend an autonomous status for higher-level

[1] Currently at: Radboud Universiteit Nijmegen, Nijmegen Institute for Cognition and Information, The Netherlands

[2] Part of this work was performed at a position at Utrecht University, Department of Philosophy, The Netherlands

[3] ‘Such identities, one could argue, are essential to the ontological simplification that we seek in theory reduction, for they enable us to dispense with facts involving F and G as something in addition to facts involving F^* and G^*.’ (…) ‘It shows that fewer basic laws, and fewer basic expressions, fully suffice for the description and explanation of the phenomena of a given domain.’ Kim (1996, p. 215)

theories in special sciences such as Biology and Cognitive Science.

Historically, a main strategy to attack eliminative tendencies in the literature on reduction has been to cast doubt on the existence of reduction relations between higher-level and lower-level theories; e.g., Davidson (1993); Bennett and Hacker (2003). The idea is that when bridge relations do not exist, it is impossible to use them for elimination of the higher-level theory; thus it is protected. Moreover, often also doubt is put forward about the existence of psychological laws to be used in explanation of human behavior.[4] These are debatable strategies for several reasons. Scientific progress made in areas such as Biochemistry and Neuroscience makes (and still will make) it harder and harder to maintain that reduction relations cannot exist. It is more constructive to investigate what can be gained for the higher-level theory from these scientific achievements rather than to feel forced to ignore them. Moreover, not (yet) having a perfect higher-level theory is not an argument for its inexistence. Finally, it can easily be questioned whether it is desirable to protect scientific theories that have no connection to lower-level theories.

Reduction Relations and How to Exploit Them

The strategy discussed in this paper accepts the existence of reduction relations, shaped in one way or the other, but claims that this can go hand in hand with (and even support) an anti-reductionist view on the development and use of higher-level theories. To clarify the different possible positions more explicitly, we distinguish between:

- reduction in a *structural sense*: as a *reduction relation*, already established or being established, between two theories and their ontologies and laws, and
- the *pragmatics* related to reduction: the *use* of an existing or to be achieved reduction relation in scientific practice.[5]

This distinction makes four positions possible: anti-reductionist or reductionist in both senses, and anti-reductionist in one sense but reductionist in the other sense. Positions discussed above include those reductionist in both senses, respectively anti-reductionist in both senses. A position that claims to be anti-reductionist in the structural sense but reductionist in the pragmatics sense ignores the higher-level theory without assuming reduction relations to a lower-level theory. There may be cases that a higher-level theory is simply so far from what is required that the lack of quality can be the reason that no reduction relations can be found, and therefore this theory just should be discarded. Eliminative reductionism (sometimes called eliminative materialism; e.g., Churchland, 1986) is such a position: it proposes to get rid of folk psychology (in favour of a neuroscientific theory to be developed), because it is not related to a lower-level theory and is not a scientific theory; see also Bennett and Hacker (2003, pp. 366-377).

The position discussed in this paper falls in the remaining category; it combines a reductionist perspective in the structural sense with an anti-reductionist perspective in the pragmatics sense. In other words, the claim is that an actual or envisioned reduction in a structural sense (reduction relations) can be useful to enforce the use and (further) development of the higher-level theory. In Jonker, Treur and Wijngaards (2002), this claim is illustrated for a case study on the development of higher-level languages in the area of Computer Science and AI. This case study shows how reduction relations between higher-level and lower-level languages within the computer, hidden for the human, support the use of the higher-level language by the human in practice, thus increasing the complexity of applications and the speed with which such applications are developed.

Higher-Level Explanations

In relation to reduction, different types of explanations can be distinguished, according to what is to be explained (explanandum) and what explains (explanans). The explanandum can be either an instance of observed behavior, the occurrence of an internal mental state, or a general law or regularity of the higher-level theory. The explanans can be found within the higher-level theory, the lower-level theory and/or the reduction relations. Different types of explanation can be distinguished; for example:

a explanation of observed behavior from the higher-level theory
b explanation of observed behavior from the lower-level theory
c explanation of the occurrence of a high-level mental state from the higher-level theory
d explanation of the occurrence of a high-level mental state from the occurrence of lower-level states
e explanation of a higher-level law or regularity from lower-level laws or regularities

Note that the level of description of the observed behavior in a) and b) corresponds to the level of the theory from which the explanation is provided. Explanations where the explanandum is from the higher-level theory and the explanans from the lower-level theory (as in d) and e)) are sometimes called reductive explanations. These are explanations of how higher-level elements work or are realised in terms of underlying (lower-level) mechanisms; see also Kim (2005, Ch. 4, pp. 93-120).

This section discusses why explanation of behavior with explanans in a higher level-theory (type a)) are considered more useful in practice than those from a lower-level theory (type b)). First a perspective developed by Jackson and

[4] '.. a neural pattern in one's brain cannot have the logical consequences of believing something: namely, being either right or wrong about what one believes. For there is no such thing as a neural configuration's being right or wrong about the truth of a proposition (…) So, if there is no sense to literally *identifying* neural states and configurations with psychological attributes, there cannot be general bridge principles linking the reducing entities (neural configurations) with the entities that are to be reduced (psychological attributes). But if there can be no bridge principles, then there is no hope for any form of reduction that will allow one to derive the laws governing phenomena at the higher level of psychology from the laws governing phenomena at the neural level. So this form of derivational reductionism is chimerical.

Not only are there no bridge principles allowing any form of ontological reduction of psychological attributes to neural configurations, but it is far from evident that there is anything that can be dignified by the name of *psychological laws* of human action, that might be reduced to, and so explained by reference to, whatever neurological laws might be discovered. For, as far as explaining human action is concerned, it is clear enough that although there are many different kinds of explanation of why people act as they do, or why a certain person acted as they did, they are not *nomological* explanations (i.e., they are nor explanations that refer to a natural *law* of human behavior).' Bennett and Hacker (2003), pp. 361-362

[5] Notice that those papers describing a position as 'eliminative' in some sense usually address this aspect of pragmatics. However, to be not committed to the more narrow scope of these approaches as present in the current literature, here the more neutral term 'pragmatics' is used.

Pettit (1988, 1990) is discussed. Subsequently, Dennett (1987)'s view on the use of intentional stance versus physical stance explanations is addressed.

Jackson and Pettit (1988, 1990) develop a notion of higher-level explanation, meant to be suitable for special sciences such as Biology, Cognitive Science and Social Sciences: *program explanation*. According to this type of explanation, 'G occurred because F occurred' (for higher-level properties F and G) can be an adequate explanation in the following way: F ensures ('programs for') some lower-level property P, which causes G. Or: F ensures ('programs for') some lower-level property P, which causes a lower-level property Q for which G is a higher-level description. For example, 'Why did the vase break?' can be explained by: 'Because it was fragile'. Here the higher-level property of being fragile ensures the lower-level property of having a specific molecular structure, and similarly for the broken state of the vase. Jackson and Pettit emphasize that such a form of higher-level explanation, based on the higher-level theory (type a)) has advantages over *causal explanation*, based on a lower-level theory (type b)), in the sense that other information is provided, which implies *increased genericity*: it not only applies to the actual world, but also to other possible worlds, and thus to, possibly, the future.[6]

As opposed to explanations from a direct physical perspective (the *physical stance*), Dennett, (1987, 1991) advocates use of the *intentional stance*; cf. Dennett (1987), pp. 37-39; Dennett (1991), pp. 37-42. He emphasizes the advantage of the explanation of behavior from a higher-level theory (type a)) compared to explanation from a lower-level theory (type b)). He considers tractability a criterion for which the higher-level explanation makes a difference.[7]

Reduction Relations

In the classical approach (following Nagel, 1961), reduction relations are based on (biconditional) bridge principles that relate the expressions in the language of a higher-level theory T_2 to expressions in the language of the lower-level theory T_1. Kim (2005, pp. 98-102) calls this *bridge law reduction*, as opposed to the type of reduction he puts forward: *functional reduction* (see also the next paragraph). Also functional reduction is based on relationships between entities in the languages of the two theories; these relationships are a bit less direct than biconditional bridge principles. A type of reduction relation not relating syntactical elements, but model structures is *structural reduction*; e.g., Balzer and Moulines (1996). Using the structuralist perspective, Bickle (1998, pp. 199-211) discusses *revisionary* (or *new wave*) *reduction*. Bickle (1998, pp. 205-208), illustrates this account for the higher-level (folk psychological) and lower-level (neurobiological) explanation in the context of Hawkins and Kandel's (1984a,b) case.[8] Here he proposes that by relating a folk-psychological explanation to a neurobiological account, a decision can be made to enrich the former by introducing some new intermediary states, based on the more detailed path provided by the latter. He puts forward the possibility to use a reduction relation in scientific practice not to eliminate the theory T_2 in favour of a theory T_1, but to extend or improve the theory T_2 on the basis of theory T_1. Therefore, abandoning explanation on the basis of T_2 and replacing such explanation by explanation on the basis of T_1 is not at issue; on the contrary, the explanatory value of T_2 is strengthened by the process of co-evolution of T_2 and T_1. Based on an extended case study on memory consolidation, he describes reduction relations and their use.[9] For further development, see Bickle (2003).

Kim (2005, pp. 98-102) puts forward *functional reduction*; see also Kim, 1998, pp. 19-23, 97-103). In schematic form (Kim, 2005, pp. 101-102):

STEP 1 [FUNCTIONALIZATION OF THE TARGET PROPERTY]

Property M to be reduced is given a functional definition of the following form:

Having M $=_{def.}$ having some property or other P (in the reduction base domain) such that P performs causal task C.

For a functionally defined property M, any property in the base domain that fits the causal specification definitive of M (that is, a property that performs causal task C) is called a "realizer" of M.

STEP 2 [IDENTIFICATION OF THE REALIZERS OF M]

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 M perform task C.'

The functionalization of mental properties makes them relational: they are specified by how they relate to other properties. Kim (1996, pp. 200-202) uses a similar idea to solve problems in the area of representational content of mental properties.[10] In particular, concentrating on the

[6] 'According to (Lewis, 1988), to explain something is to provide information on its causal history . . . A program explanation provides a different sort of information . . . A program account tells us what the history might have been. It gives modal information about the history, telling us for example that in any relevantly similar situation, as in the original situation itself, the fact that some atoms are decaying means that there will be a property realized - that involving the decay of such and such particular atoms - which is sufficient in the circumstances to produce radiation. In the actual world it was this, that and the other atom which decayed and led to radiation, but in possible worlds where their place is taken by other atoms, the radiation still occurs. ' (Jackson and Pettit, 1990), p. 117.

[7] '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.

[8] 'Of course, the functional profiles assigned to cognitive states on Hawkin and Kandel's neurobiological account are much more fine-grained and detailed, for that account recognizes distinctions and connections that folk psychology either lumps together or leaves extremely vague (…) we can expect that injection of some neurobiological details back into folk psychology would fruitfully enrich the latter, and thus allow development of a more fine-grained folk-psychological account that better matches the detailed functional profiles that neurobiology assigns to its representational states. There is no principled reason against such enrichment.' (Bickle, 1998), p. 207-208

[9] 'In particular, entities characterised on the reduced theory *primarily by their functional (input-output) features* get linked to complex structures (sequences and combinations of entities and processes) whose dynamics and interactions specifiable entirely within the reducing theory apply to roughly the same intended set of real-world systems, and provide causal mechanisms that explain the former's functional (input-output) profile. This, in rough outline, is the metascientific concept of a theoretical posit becoming "structured through reduction".' Bickle (2003, pp. 98-99)

[10] '… to consider beliefs to be wholly internal to the subjects who have them but consider their contents as giving *relational specifications* of the beliefs. On this view, beliefs may be neural states or other types of physical states of organisms and systems to which they are attributed. Contents, then, are viewed as ways of specifying these inner states; wide contents, then, are specifications in terms of, or under the constraints of, factors and conditions external to the subject, both physical and social, both current and historical.'
(Kim, 1996, pp. 200-201); italics in the original.

temporal dimension, a temporal relational specification can be viewed as the specification of temporal relationships of a mental state property to other patterns ('factors and conditions') in past and future. In Kim's proposal a mental state property of a subject itself is distinguished from its relationships to other items. This contrasts to some other approaches where the mental state property is considered to be ontologically constituted as one entity comprising both the subject and the related items, or where the mental state property itself is considered to be the relation between the subject and the other items (cf. Kim, 1996, pp. 200-202).

The main difference between the specification of the causal task of a mental property M and its representational content is that the former describes the functional role of M as a mediator between its close causal neighbors, whereas the latter describes possibly more complex relationships of M with states further away in position and time. The former type of specification of M is best suited for a reduction relation, for M to be mapped onto lower level properties and their causal relationships within the lower-level theory, as Kim (2005) proposes. The latter type of specification (of representational content) can be used to specify its meaning, and as such gives more detail and grounding of concepts within the higher-level theory and within reality.

Methodological Perspective

In our previous work, a number of cases studies already have been undertaken to explore the usefulness in practice of Kim's relational perspective on reduction and representational content. In these case studies, different types of mental states have been examined: beliefs, desires and intentions, trust, adaptation, core consciousness, extended mind. It was shown how for practical (formal) analysis and simulation the less complex higher-level theory can be used. Based on this amount of work, a first conclusion is that this perspective works fine in practical contexts. Contrary to Kim's quotes above, this perspective is not used to eliminate the higher-level description, but instead to strengthen it, giving it a solid basis. Based on these experiences and the issues described above, the methodological perspective is formulated as follows.

To analyse a certain high-level state property, perform the following steps:

1. analyse in a relational manner the functional role of the state property in relation to other properties and formalise *relational specifications for the causal roles* involved
2. distinguish a number of behaviors in which the property fulfills a role, and formally specify these behaviors from an external perspective
3. analyse the property in a relational manner and formulate formalised relational specifications for its representational content
4. analyse formally whether the specifications in 1., 2., and 3. are coherent, i.e., whether they relate correctly; if needed correct them
5. analyse the property from a neurological point of view by identifying neurological states that perform the same causal role; if needed improve the specifications in 1., 2., and 3.
6. show how different types of explanations can be distinguished as used in practice for different purposes
   a. explanation of behavior
      i. using laws or regularities from the higher-level theory, independent of reduction relations
      ii. using the lower-level theory
   b. explanation of the occurrence of a high-level (mental) state
      i. using the higher-level theory, relating it over time to other states
      ii. using reduction relations, relating it to the occurrence of lower-level states
   c. explanation of the mechanisms behind a higher-level law or regularity:
      i. using reduction relations, relating it to lower-level laws or regularities

Most of these items can be supported by formal languages and supporting tools. For 1. causal modelling formats can be used, such as the LEADSTO language; cf. (Bosse et al., 2005). For 2. and 3. a more expressive language is needed such as TTL, or a temporal language as CTL; cf. (Bosse et al., 2006; Goldblatt, 1992). For 4. the TTL checking environment can be used and model checking tools such as SMV[11]; cf. (McMillan, 1993). For 5. causal formats or LEADSTO can be used, for 6. the TTL environment.

Case Study

The methodology is illustrated for a case study: the adaptive behavior of *Aplysia*. *Aplysia* is a sea hare that is often used to do experiments. It is an interesting case to use as illustration, since its internal neural mechanisms are relatively simple, and therefore well understood. *Aplysia* is able to learn: it performs classical conditioning in the following manner. This (a bit simplified) description is mainly based on (Gleitman, 1999), pp. 155-156. Initially the following behavior is shown: a tail shock leads to a response (contraction), and a light touch on its siphon is insufficient to trigger such a response. After the subject is a number of times touched lightly on its siphon and then shocked on its tail (as a consequence it responds), it turns out that the behavior has changed: the animal also responds (contracts) on a siphon touch. In this section some indications are given on how to address this example (in particular, the high-level state property of 'having learnt to respond to a siphon touch') from the perspective described above.

1. Relational specification of states and functional roles

Below, a formal model is provided of the states and processes involved in *Aplysia*'s learning behavior. The basic building blocks of this model are *state properties* (descriptions of states of the process at a certain time point) and their functional roles expressed by *executable properties* (causal relations between state properties at different time points). The state properties used are:

tail_shock	the animal is shocked on its tail
siphon_touch	the animal is touched on its siphon
contraction	the animal contracts
sr(siphon_touch)	sensory representation of a siphon touch
sr(tail_shock)	sensory representation of a tail shock
prep(contraction)	preparation state for contraction
c1(r1)	control state for effect of sr(siphon_touch) on preparation
c2(r2)	control state for effect of sr(tail_shock) on preparation

In addition, the following executable properties are identified to describe the functional roles of the state properties. Here, the expression $\alpha \bullet\!\!\twoheadrightarrow \beta$ (pronounced α *leads to* β) informally means the following: if state property

11 http://www.cs.cmu.edu/~modelcheck/smv.html

α holds for a certain time interval, then after some delay, state property β will hold for a certain time interval. For a precise definition, see (Bosse et al., 2005).

LP1 $\text{tail\_shock} \bullet\!\!\rightarrow \text{sr(tail\_shock)}$

LP2 $\text{siphon\_touch} \bullet\!\!\rightarrow \text{sr(siphon\_touch)}$

LP3 $\text{sr(tail\_shock)} \wedge \text{sr(siphon\_touch)} \wedge \text{c1(r1)} \wedge \text{r1<0.75} \bullet\!\!\rightarrow \text{c1(r1+0.25)}$

LP4 $\text{sr(siphon\_touch)} \wedge \text{c1(r1)} \wedge \text{r1>0.6} \bullet\!\!\rightarrow \text{prep(contraction)}$

LP5 $\text{sr(tail\_shock)} \wedge \text{c2(r2)} \wedge \text{r2>0.05} \bullet\!\!\rightarrow \text{prep(contraction)}$

LP6 $\text{prep(contraction)} \bullet\!\!\rightarrow \text{contraction}$

LP7 $\text{start} \bullet\!\!\rightarrow \text{c1(0.0)} \wedge \text{c2(0.1)}$

The relational specification of the functional role of an internal state property concerns relationships both backward and forward in time. Given the model provided above, when looking backward, the functional role of, e.g., state property c1(0.75) (which represents the fact that the animal has learnt to respond to a siphon touch) is the causal relationship between c1(0.75) and the (past) states that cause c1(0.75). Thus, looking backward the functional role of c1(0.75) is described by executable property LP3. Likewise, when looking forward the functional role of c1(0.75) is described by the causal relationship between c1(0.75) and the (future) states that are caused by c1(0.75), i.e., by executable property LP4.

2. Related behavior specifications

In this step, various behaviors can be described in which *Aplysia*'s adaptivity fulfills a role. In general, example types of behavior that can be used in this step are high-level cognitive functions such as reasoning, planning and language processing. With respect to the above case study, an example behavior, specified from an external perspective, is the following: "if a siphon touch occurs, and at three different earlier time points t1, t2, t3, a siphon touch occurred, which was directly followed by a tail shock, then the animal contracts". To make this expression more precise, it can be formalized in a temporal language. For example, in the Temporal Trace Language (TTL) by (Bosse et al., 2006), the expression is formalized as follows:

GP1 ≡

$$\begin{array}{l}
\forall t\ \text{state}(\gamma, t) \models \text{siphon_touch}\ \& \\
\exists t1, t2, t3, u1, u2, u3\ \ t1 < u1 < t2 < u2 < t3 < u3 < t\ \& \\
\text{state}(\gamma, t1) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u1) \models \text{tail_shock}\ \& \\
\text{state}(\gamma, t2) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u2) \models \text{tail_shock}\ \& \\
\text{state}(\gamma, t3) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u3) \models \text{tail_shock} \\
\quad\quad \Rightarrow\ \exists t' \geq t\ \ \text{state}(\gamma, t') \models \text{contraction}
\end{array}$$

3. Relational specification of representational content

The relational specification of the representational content of an internal state property can be addressed both backward and forward in time. When looking backward, the backward representational content of an internal state can be described, for example, by relating it to a history of past world states. The representational content of state property c1(0.75) can be described informally as follows: 'if at three different earlier time points t1, t2, t3, a siphon touch occurred, which was directly followed by a tail shock, then c1(0.75) will occur', and conversely. In TTL, the expression is formalized as follows:

$$\begin{array}{l}
\forall t1, t2, t3, u1, u2, u3\ \ t1 < u1 < t2 < u2 < t3 < u3\ \& \\
\text{state}(\gamma, t1) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u1) \models \text{tail_shock}\ \& \\
\text{state}(\gamma, t2) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u2) \models \text{tail_shock}\ \& \\
\text{state}(\gamma, t3) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u3) \models \text{tail_shock} \\
\quad\quad \Rightarrow\ \exists t > u3\ \ \text{state}(\gamma, t) \models \text{c1(0.75)}\ \& \\
\forall t\ \text{state}(\gamma, t) \models \text{c1(0.75)} \Rightarrow \\
\exists t1, t2, t3, u1, u2, u3\ \ t1 < u1 < t2 < u2 < t3 < u3 < t\ \& \\
\text{state}(\gamma, t1) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u1) \models \text{tail_shock}\ \& \\
\text{state}(\gamma, t2) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u2) \models \text{tail_shock}\ \& \\
\text{state}(\gamma, t3) \models \text{siphon_touch}\ \&\ \text{state}(\gamma, u3) \models \text{tail_shock}
\end{array}$$

Similarly, when looking forward, the representational content of an internal state can be described by relating it to future world states. The future representational content of state property c1(0.75) can be informally described as follows: 'if c1(0.75) occurs, then whenever the animal is touched on its siphon, it will contract', and conversely. In the TTL language, the expression is formalized as follows:

$$\begin{array}{l}
\forall t\ \text{state}(\gamma, t) \models \text{c1(0.75)} \Rightarrow \\
\quad [\ \forall t' > t\ \text{state}(\gamma, t') \models \text{siphon_touch} \Rightarrow \exists t'' > t'\ \text{state}(\gamma, t'') \models \text{contraction}\ ] \\
\quad\quad \&\ [\forall t\ [\ \forall t' > t\ \text{state}(\gamma, t') \models \text{siphon_touch} \Rightarrow \\
\quad\quad\quad\quad \exists t'' > t'\ \text{state}(\gamma, t'') \models \text{contraction}\ ] \\
\quad\quad\quad \&\ \ \exists t' > t\ \text{state}(\gamma, t') \models \text{siphon_touch}\ ] \Rightarrow\ \text{state}(\gamma, t) \models \text{c1(0.75)}
\end{array}$$

4. Analysis of the coherency of the specifications in 1-3.

In general, if specifications are sufficiently complete and correct, the specifications found in 1. (maybe extended for a number of other related internal states) should logically entail those found in 2. and those found in 3. This can be verified using methods for logical and formal analysis. If discrepancies are found, improvements are to be made. In terms of the case study, part of the analysis has been to verify (using the SMV environment; cf. McMillan, 1993) that indeed the local properties LP1 through LP9 together entail global property GP1. Likewise, it has been verified that these local properties together entail the representational content specifications.

5. Analysis of the state from a neurological point of view

To analyze an internal state property from a neurological point of view, neurological states have to be identified that perform the same causal role. In (Gleitman, 1999) the internal neural mechanisms for *Aplysia*'s conditioning are described (see also Figure 1). A tail shock activates a sensory neuron SN1. Activation of this neuron SN1 activates the motoneuron MN; activation of MN makes the sea hare move. Moreover, a siphon touch activates the sensory neuron SN2. Activation of this sensory neuron SN2 normally does not have sufficient impact on MN to activate MN. After learning, activation of SN2 has sufficient impact to activate MN. In addition, activation of SN1 also leads to activation of the intermediary neuron IN. If both SN2 and IN are activated simultaneously, this changes the state of the synapse between SN2 and MN: it causes a state of this synapse in which it produces more neurotransmitter if SN2 is activated. As a result, after a number of trials, activation of SN2 also yields activation of MN.

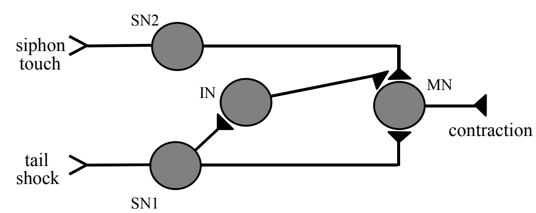

Figure 1: *Aplysia*'s internal neural mechanisms.

Such neurological states can be related to the functional states as described in step 1. by Kim (2005)'s functional reduction. Some example relations are the following:

- sr(tail_shock) relates to SN1
- sr(siphon_touch) relates to SN2
- c1(0.0) relates to a synapse state forming a weak connection between SN2 and MN
- c1(0.75) relates to a synapse state forming a strong connection between SN2 and MN
- prep(contraction) relates to MN

Note that IN is not used in these relations. This shows that the lower-level theory gives a more fine-grained account than the higher-level theory.

6. Different types of explanations for different purposes
The knowledge generated under 5. can be considered as a focus on specialized neurological knowledge, that can be used to verify the specifications from a different focus in 1. to 3. For the different types of explanations discussed above (i.e., type a), b) and c)), instances can be formulated based on the specifications obtained in 1. to 3. For example, an explanation of type a) explains at a higher level why an animal contracts in certain situations. The explanation of this behavior of type a) refers to higher-level concepts such as sensory representations and preparation states.

Discussion

Within philosophical literature, especially in the area of Cognitive Science and Philosophy of Mind, higher-level cognitive concepts are often questioned. Neurological and biochemical mechanisms underlying human behavior provide alternative explanations of action. Not seldom it is argued that reduction relations between cognitive states and neurophysiologic states show that higher-level cognitive concepts can be eliminated and replaced by neurophysiological concepts. To clarify the different possible positions more explicitly, in this paper a distinction is made between reduction in a structural sense (as a reduction relation between two theories and their ontologies and laws) and the pragmatics related to reduction (the use of a reduction relation in scientific practice). This distinction makes four positions possible: anti-reductionist or reductionist in both senses, and anti-reductionist in one sense but reductionist in the other sense.

The position discussed here in some detail combines a reductionist perspective in the structural sense with an anti-reductionist perspective in the pragmatics sense. In other words, the claim is that an actual or envisioned reduction in a structural sense (reduction relations) can be useful to enforce the use and (further) development of the higher-level theory. A practical methodology is proposed to develop higher-level theories by grounding them in lower-level theories. The methodology incorporates Kim (1996, 2005)'s perspective on relational specification and functional reduction, and provides technical support by formal methods and tools as developed in Computer Science and Artificial Intelligence. It was shown how this methodology can be used to formalise the adaptive behaviour of *Aplysia*. Another case study that has been undertaken addresses the processes leading to core consciousness according to Damasio (2000) (see URL http://www.cs.vu.nl/~tbosse/reduction/consciousness.doc).

References

Balzer, W., and Moulines, C.U. (1996). *Structuralist Theory of Science*. Walter de Gruyter, Berlin.

Bennett, M. R., and Hacker, P. M. S. (2003). Philosophical foundations of neuroscience. Malden, MA: Blackwell.

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

Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthless Reductive Account*. Kluwer Academic Publishers.

Bosse, T., Jonker, C.M., Meij, L. van der, Sharpanskykh, A., and Treur, J. (2006). A Temporal Trace Language for the Formal Analysis of Dynamic Properties. Technical Report, Vrije Universiteit Amsterdam. http://www.few.vu.nl/~treur/TTL.pdf

Bosse, T., Jonker, C.M., Meij, L. van der, and Treur, J. (2005). LEADSTO: a Language and Environment for Analysis of Dynamics by SimulaTiOn. In: Eymann, T. et al. (eds.), *Proceedings of the Third German Conference on Multi-Agent System Technologies, MATES'05*. Lecture Notes in AI, vol. 3550. Springer Verlag, pp. 165-178.

Churchland, P.S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. MIT Press, Cambridge, MA.

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

Davidson, D. (1993). *Thinking Causes*. In: J. Heil & A. Mele (eds)., *Mental Causation. Clarendon Press, Oxford.*

Dennett, D.C. (1987). *The Intentional Stance.* MIT Press. Cambridge Mass.

Dennett, D.C. (1991). Real Patterns. *The Journal of Philosophy*, vol. 88, pp. 27-51.

Gleitman, H. (1999). *Psychology*. W.W. Norton & Company, New York.

Goldblatt, R. (1992). *Logics of Time and Computation*, 2nd edition, CSLI Lecture Notes 7.

Hawkins, R.D., and Kandel, E.R. (1984a). Is There a Cell-Biological Alphabet for Simple Forms of Learning? *Psychological Review*, vol. 91, pp. 375-391

Hawkins, R.D., and Kandel, E.R. (1984b). Steps Toward a Cell-Biological Alphabet for Elementary Forms of Learning. In: G. Lynch, J.L. McGaugh, and N.M. Weinberger (eds.), *Neurobiology of Learning and Memory*, Guilford Press, New York, pp. 385-404, vol. 91, pp. 375-391

Jackson, F., and Pettit, P. (1988). Functionalism and Broad Content. *Mind*, vol. 97, pp. 381-400.

Jackson, F., and Pettit, P. (1990). Program Explanation: A General Perspective. *Analysis*, vol. 50, pp. 107-117.

Jonker, C.M., Treur, J., and Wijngaards, W.C.A. (2002). Reductionist and Antireductionist Perspectives on Dynamics. *Philosophical Psychology Journal*, vol. 15, pp. 381-409.

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

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

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

Lewis, D. (1988). Causal Explanation. In: *Philosophical Papers*, vol. 2. Oxford: Basil Blackwell.

McMillan, K.L., (1993). Symbolic Model Checking: An Approach to the State Explosion Problem. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1992. Published by Kluwer Academic Publishers, 1993.

Nagel, E. (1961). The Structure of Science. London: Routledge & Kegan Paul.

Smart, J.J.C. (1959). Sensations and Brain Processes. *Philosophical Review*, vol. 68, pp. 141-156. Reprinted in Rosenthal, D.M. (ed.), *The Nature of Mind*. Oxford University Press, Oxford, 1991.

Wegner, D.M. (2002). *The illusion of conscious will*. Cambridge, MA. MIT Press.