Cognitive and Biological Agent Models for Emotion Reading

Zulfiqar A. Memon, Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence

De Boelelaan 1081, 1081 HV Amsterdam, the Netherlands

http://www.few.vu.nl/~{zamemon, treur} {zamemon, treur}@few.vu.nl

Abstract

This paper focuses on how capabilities to interpret another agent's emotions, and their biological realisation can be modelled. First a cognitive and a biological agent model to generate emotional states are introduced. These models involve the generation and sensing of body states. Next it is shown how emotion reading can be modelled both at a cognitive and at a biological level, following the Simulation Theory approach to mindreading. Furthermore, a cognitive Theory Theory model for emotion reading is presented. Finally, it is shown how the cognitive and biological agent models can be related.

1. Introduction

From an evolutionary perspective, mindreading in humans (and some other kinds of animals) has developed for a number of aspects, for example, intention, attention, emotion, knowing; e.g., [1], [4], [6]. Two philosophical perspectives on having a Theory of Mind are Simulation Theory and Theory Theory. In the first perspective it is assumed that mindreading takes place by using the facilities involving the own cognitive states that are counterparts of the cognitive states attributed to the other person. For example, the state of feeling pain oneself is used in the process to determine whether the other person has pain. The second perspective is based on reasoning using knowledge about relationships between cognitive states and observed behaviour. An example of such a pattern is: 'I hear that she says 'ouch!'. Having pain causes saying 'ouch!'. Therefore she has pain'.

In [4] mind reading models from a Theory Theory perspective have been described for an agent reasoning about another agent's intentions and actions. For reasoning processes in general it is not clear how they relate to underlying biological processes. In particular, this holds for the Theory Theory perspective on mindreading. The Simulation Theory perspective may seem to have a more direct relationship to the person's own biological structures for a certain cognitive state. The biological structures for a person's own emotions are often considered to have a causal/ temporal direction from the emotional state to the (bodily) expression of the emotion, whereas a mindreading process proceeds the other way around: starting from the expression, to determine the underlying emotion. So, a puzzling question is how it is possible to use simulation for a process that, temporally has the opposite direction. As part of the work reported in this paper this paradox was resolved.

As part of a solution a generalisation of Damasio's concept of 'body loop' as put forward in [5] and formalised in [3], was used. This perspective distinguishes the (bodily) emotional response to a stimulus from feeling the emotion (sometimes called the emotional feeling), which is caused by sensing the own bodily response. An extension of this idea can be obtained by assuming that the body loop is not processed once, as a linear causal chain starting with the stimulus and via the body ending up in the feeling, as assumed in [3], [5]. Instead a recursive body loop can be assumed: a converging positive feedback loop based on reciprocal causation between emotional state (gradually more feeling) and body state (gradually stronger expression). Triggered by a stimulus, after a number of rounds this loop ends up in an equilibrium for both states. It turns out that such a recursive body loop model for emotions is an appropriate basis to obtain emotion reading models from the Simulation Theory perspective.

In the paper, Sections 2 and 3 show both a cognitive and biological agent model for emotion generation based on a recursive body loop. Section 4 shows how this can be (re)used to model emotion reading at a cognitive level from the Simulation Theory perspective. Furthermore, in Section 5 it is discussed how emotion reading can be modelled at the biological level. Next, in Section 6 the cognitive and biological agent models for emotion generation and emotion reading are related to each other. Finally in Section 7 a cognitive agent model for emotion reading from the Theory Theory perspective is introduced and related to the other models.

2. Emotion generation: a cognitive model

In this section the model *EGCM* (Emotion Generation Cognitive Model) to generate emotional states for a given stimulus is introduced. It adopts from [5] the idea of a 'body loop' and 'as if body loop', but extends this by making these loops recursive. According to the original idea, a body loop roughly proceeds according to the following causal chain; see [3], [5]: sensing a stimulus → sensory representation of stimulus → (preparation for) bodily response → sensing the bodily response → sensory representation of the bodily response → feeling the emotion. In the model introduced here an essential addition is that the body loop (or as if body loop) is extended to a *recursive body loop* (or *recursive as if body loop*) by assuming that the preparation of the bodily response is also affected by the state of feeling the emotion (also called emotional feeling): feeling the emotion → preparation for bodily response as an additional causal relation. Thus the obtained model is based on reciprocal causation relations between emotional feeling and body states. Moreover, both the bodily response and the emotional feeling are assigned a level or gradation, expressed by a number, which is assumed dynamic. The causal cycle is modelled as a positive feedback loop, triggered by the stimulus and converging to a certain level of emotional feeling and body state. Here in each round of the cycle the next body state has a level that is affected by both the levels of the stimulus and the emotional feeling state, and the next level of the emotional feeling is based on the level of the body state. In the more detailed model described below, the combined effect of the levels of the stimulus and the emotional state on the body state is modelled as a weighted sum with equal weights 0.5.

In the description of the detailed cognitive model the temporal relation a ↠ b denotes that when a state property a occurs, then after a certain time delay (which for each relation instance can be specified as any positive real number), state property b will occur. In this language (called LEADSTO) both logical and numerical calculations can be specified, and a dedicated software environment is available to support specification and simulation; for more details see [2].

The specification (both informally and formally) of the model for emotion generation based on a recursive body loop is as follows.

LP1 Sensing a stimulus
If stimulus s occurs then a sensor state for s will occur.
s ↠ sensor_state(s)

LP2 Generating a sensory representation of a stimulus
If a sensor state for s occurs, then a sensory representation for s will occur.
sensor_state(s) ↠ srs(s)

LP3 From sensory representation and emotion to preparation
If a sensory representation for s occurs and emotion e has level v, then preparation state for face expression f will occur with level (1+v)/2.
srs(s) & emotion(e, v) ↠ preparation_state(f, (1+v)/2)

LP4 From preparation to body modification
If preparation state for face expression f occurs with level v, then the face is modified to express f with level v.
preparation_state(f, v) ↠ effector_state(f, v)

LP5 From body modification to modified body
If the face is modified to express f with level v, then the face will have expression f with level v.
effector_state(f, v) ↠ own_face(f, v)

LP6 Sensing a body state
If facial expression f with level v occurs, then this facial expression is sensed.
own_face(f, v) ↠ sensor_state(f, v)

LP7 Generating a sensory representation of a body state
If facial expression f of level v is sensed, then a sensory representation for face expression f with level v will occur.
sensor_state(f, v) ↠ srs(f, v)

LP8 From sensory representation of body state to emotion
If a sensory representation for face expression f with level v occurs, then emotion e is felt with level v.
srs(f, v) ↠ emotion(e, v)

LP9 Imputation
If a certain emotion e is felt, with level ≥ 0.5 and a sensory representation for s occurs, then emotion e will imputed to s.
srs(s) & emotion(e, v) & v≥0.5 ↠ imputation(s, e)

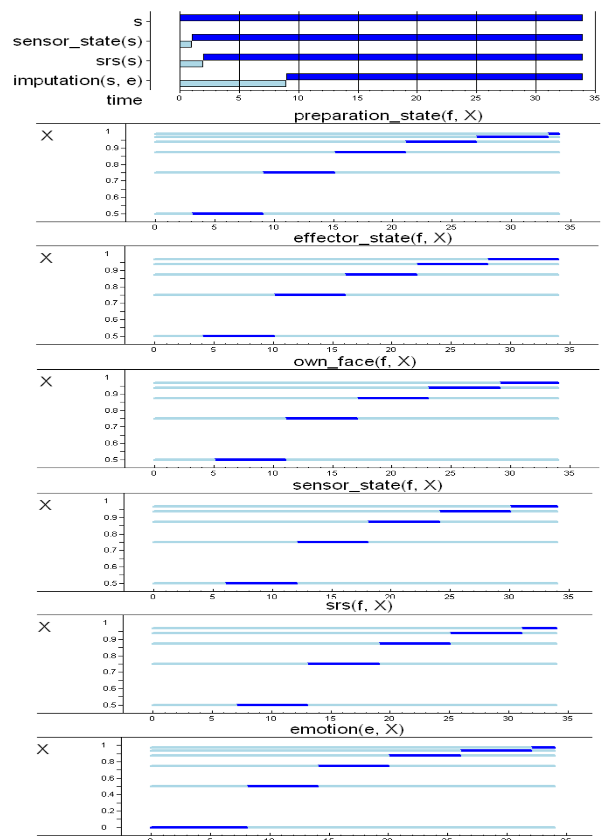

Figure 1 Simulation of cognitive model EGCM

In the imputation state, the emotion e felt, is related to the stimulus s, that triggered the emotion generation process. Note that this state makes sense in general, for any type of stimulus s, as usually a person does not only feel an emotion, but also has an awareness of what causes an emotion; what in [5] is called a state of conscious feeling also plays this role.

Based on the model, a number of simulations have been performed; for an example, see Figure 1 (here the time delays within the temporal LEADSTO relations were taken 1 time unit). In this figure, where time is on the horizontal axis, the upper part shows the time periods in which the binary logical state properties s, sensor_state(s), srs(s), imputation(s, e) hold (indicated by the dark lines): respectively from time point 0, 1, 2 and 9. Below this part for the other state properties values for the different time periods are shown (by the dark lines). For example, the preparation state for f has value 0.5 at time point 3, which is increased to 0.75 and further at time point 9 and further. The graphs show how the recursive body loop approximates converging states both for emotion and face expression: value 1 for both.

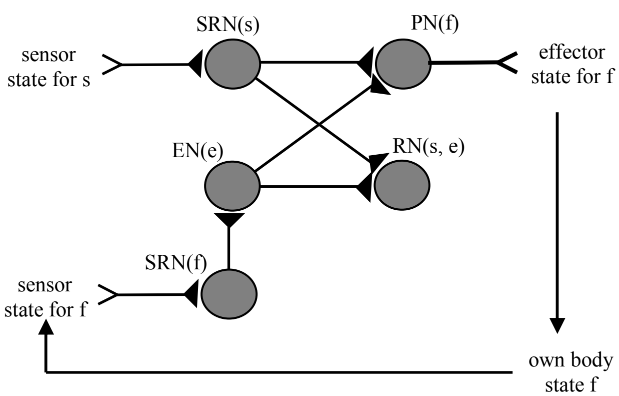

Figure 2 Overview of the biological model EGBM

3. Emotion generation: a biological model

As a next step, the biological model *EGBM* (Emotion Generation Biological Model) for the emotion generation process based on a recursive body loop is described. This biological model refers to activation states of (groups of) neurons and the body. An overall picture of the connection for this model is shown in Figure 2. Here each node stands for a group of one or more neurons. The nodes can be interpreted as follows:

SRN(s) *sensory representation neuron(s) for* s
SRN(f) *sensory representation neuron(s) for* f
PN(f) *preparation neuron(s) for own body state* f
EN(e) *neuron(s) for felt emotion* e
RN(s, e) *neuron(s) for representation that* s *carries emotion* e

The connections with their strengths are specified by:

connectedto(s, sensor_state(s), 1)
connectedto(sensor_state(s), SRN(s), 1)
connectedto(EN(e), SRN(s), PN(f), 0.5, 0.5)
connectedto(PN(f), effector_state(f), 1)
connectedto(effector_state(f), own_face(f), 1)
connectedto(own_face(f), sensor_state(f), 1)
connectedto(sensor_state(f), SRN(f), 1)
connectedto(SRN(f), EN(e), 1)
connectedto(EN(e), SRN(s), RN(s, e), 0.5, 0.5)

The generic propagation rules for activation levels used are specified in LEADSTO format as:

connectedto(X, Y, α) & activated(X, v) $\twoheadrightarrow$ activated(Y, α* v)
connectedto(X1, X2, Y, α, β) & activated(X1, v1) & activated(X2, v2) $\twoheadrightarrow$ activated(Y, α* v1+β *v2)

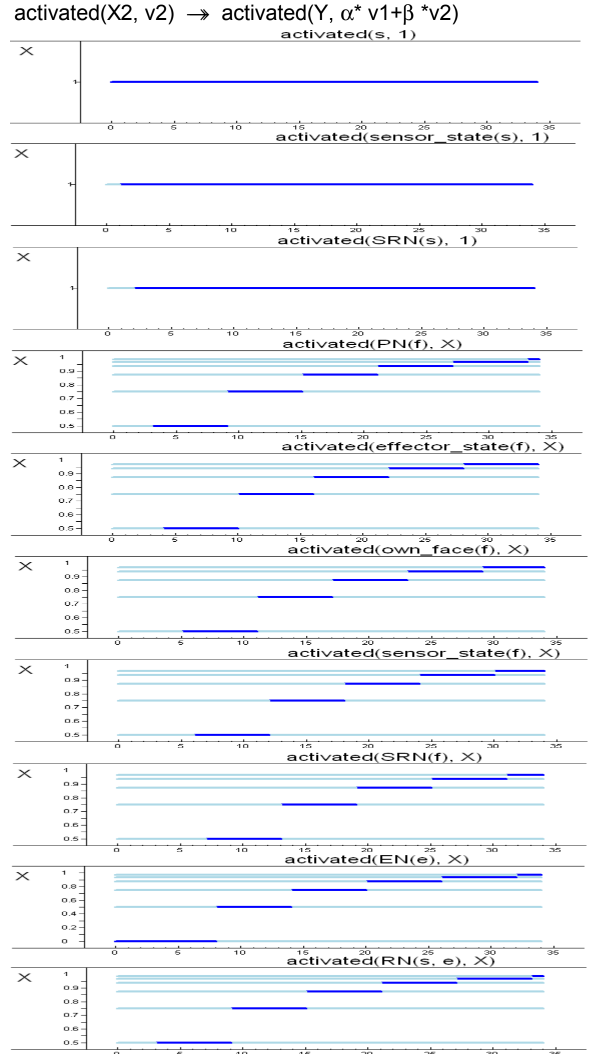

Figure 3 Example trace for biological model EGBM

These temporal relations specify that propagation of activation levels takes place by multiplying them by the strength of the connection; for input from multiple connections they are added.

Based on the biological model specifications, within the LEADSTO environment simulation traces have been generated, for example, as shown in Fig. 3. Time delays within the above two temporal LEADSTO relations again were taken 1 time unit). The graphs show the values of the various activation levels over time. Here it is shown that the recursive body loop results in an approximation of convergent activation levels for the states that relate to the emotion and the body state, among others.

4. Emotion reading: a cognitive model

A cognitive model for emotion reading for the Simulation Theory perspective should essentially be based on a model to generate the own emotions. Indeed, the model presented in the previous section can be specialised in a rather straightforward manner to the model *STCM* (Simulation Theory Cognitive Model) enabling emotion reading. The main step is that the stimulus s that triggers the emotional process, which until now was left open, is instantiated with the body state of another person, to make it specific, a face expression f of another person is considered: s = othersface(f). Indeed there is strong evidence that (already from an age of 1 hour) sensing somebody else's face expression leads (within about 300 milliseconds) to preparing for and showing the same face expression ([6], pp. 129-130). Furthermore, for the sake of illustration, following the emotion imputation, a communication about it is prepared and performed. This extension is not essential for the emotion reading capability, but shows an example of behavior based on emotion reading.

LP10 Communication preparation
If emotion e is imputed to s, then a related communication is prepared
imputation(e, s) →→ preparation_state(say(your emotion is e))

LP11 Communication
If a communication is prepared, then this communication will be performed.
preparation_state(say(your emotion is e)) →→
effector_state(say(your emotion is e))

The model described in Section 2 has been extended by the above two temporal relations in LEADSTO format, and used for simulation. An example simulation trace was obtained that for a large part coincides with the one shown in Figure 1 (with the other person's face expression f as the stimulus), with an extension as shown in Figure 4. Here also the time delays within the additional temporal LEADSTO relations were taken one time unit.

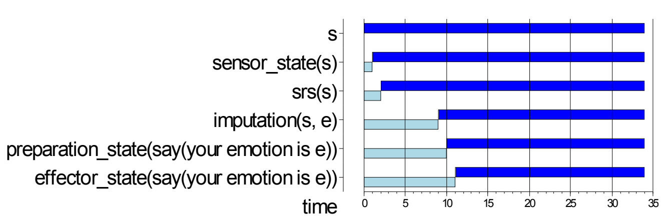

Figure 4 Trace extension: emotion reading by STCM

5. Emotion reading: a biological model

Just like in the cognitive case, a biological model *STBM* for emotion reading for the Simulation Theory perspective should essentially be based on a biological model to generate the own emotions. The biological model *EGBM* indeed can be specialised directly to enable emotion reading by taking the other person's face as the stimulus s that triggers the emotional process. A simulation trace for emotion reading is obtained of the form shown in Figure 3 with s = othersface(f); this trace can be extended with a communication part as shown in Figure 5, based on additional connections:

connectedto(RN(s, e), PN(say(your emotion is e)), 1)
connectedto(PN(say(your emotion is e)),
effector_state(say(your emotion is e)), 1)

Note that at time point 3 the neuron RN(s, e) has activation level 0.5, which is not considered high enough to count as an indication of imputation. However, after time point 9 it gets an activation level of 0.75. This is considered an appropriate indication for an imputation. This interpretation issue is made more precise in the section below about interpretation mappings between the models.

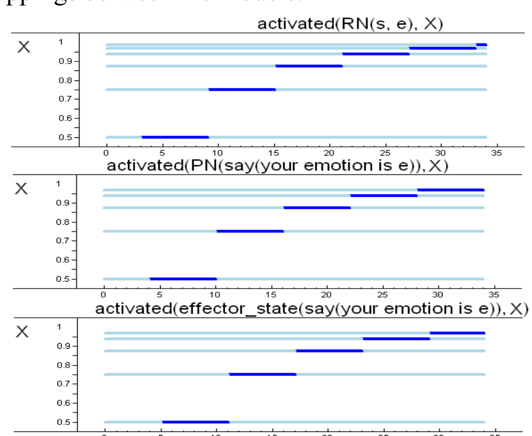

Figure 5 Trace extension: emotion reading by STBM

6. Relating the models

For cognitive models it is an interesting challenge to find out how they can be related to a biological realisation. Work on this area (of reduction) can be found in a wide variety of publications in the philosophical literature; see, for example [7]. A specific reduction approach provides a particular *reduction relation*: a way in which each cognitive property *a* can be related to a biological property *b*; this *b* is often called a *realiser* for *a*. Reduction approaches differ in how these relations are defined.

One notion to define reduction relations is a *(relative) interpretation mapping:* e.g. [9] pp. 61-65.

This is a mapping φ relating cognitive concepts a to biological concepts b, in the sense that $b = \varphi(a)$. Such a mapping is an interpretation mapping when it satisfies the property that if L is a cognitive law, then the statement $\varphi(L)$ can be derived from biological laws. Usually the mapping is assumed compositional with respect to connectives:

$$\varphi(A1 \ \& \ A2) = \varphi(A1) \ \& \ \varphi(A2)$$
$$\varphi(A1 \twoheadrightarrow A2) = \varphi(A1) \twoheadrightarrow \varphi(A2)$$

and so on. In this section interpretation mappings are applied to relate the cognitive and biological models *EGCM, STCM, EGBM, STBM*.

To define an interpretation mapping from the cognitive model *EGCM* to the biological model *EGBM* for emotion generation, a main issue to be settled is which biological states are to be interpreted as feeling the emotion, and which as the imputation of the emotional feeling to a stimulus. The interpretation mapping φ_1 from *EGCM* to *EGBM* has been defined as follows, where a criterion for considering RN(s, e) as imputation is defined by a threshold of 0.75.

φ_1(s) = activated(s, 1)
φ_1(sensor_state(s)) = activated(sensor_state(s), 1)
φ_1(srs(s)) = activated(SRN(s), 1)
φ_1(preparation_state(f, v)) = activated(PN(f), v)
φ_1(emotion(e, v)) = activated(EN(e), v)
φ_1(effector_state(f, v)) = effector_state(f, v)
φ_1(sensor_state(f, v)) = sensor_state(f, v)
φ_1(srs(f, v)) = activated(SRN(f), v)
φ_1(imputation(s, e)) = ∃v v≥0.75 & activated(RN(s, e), v)

The mapping is extended to more complex (temporal) expressions in a compositional manner:

φ_1(A1 & A2) = φ_1(A1) & φ_1(A2)
φ_1(A1 → A2) = φ_1(A1) → φ_1(A2)

The mapping maps the cognitive temporal relationships between the different state properties specified in the cognitive model *EGCM* to biological relationships between state properties entailed by the biological model *EGBM*. For example, if L is the relationship

srs(s) & emotion(e, v) ↠ preparation_state(f, (1+v)/2)

which holds in *EGCM*, then L is mapped by φ_1 onto

φ_1(L) = φ_1(srs(s) & emotion (e, v) ↠
preparation_state(f, (1+v)/2))
= φ_1(srs(s) & emotion(e, v)) ↠
φ_1(preparation_state(f, (1+v)/2)))
= φ_1(srs(s)) & φ_1(emotion(e, v)) ↠
φ_1(preparation_state(f, (1+v)/2)))
= activated(SRN(s, 1) & activated(EN(e), v) ↠
activated(PN(f), (1+v)/2)

The latter expression is not literally part of the biological model, but is entailed by it, in particular by

connectedto(EN(e), SRN(othersface(f)), PN(f), 0.5, 0.5)

together with the general rule that specifies propagation of activation through connections. In a similar way the property can be mapped that expresses that an emotion is imputed to a sensed stimulus: the temporal relation L' given by

srs(s) ↠ imputation(s, e)

is entailed by the temporal relations in the model *EGCM*. It is mapped as follows:

φ_1(L') = φ_1(srs(s) ↠ imputation(s, e))
= φ_1(srs(s)) ↠ φ_1(imputation(s, e))
= activated(SRN(s), 1) ↠ ∃v v≥0.75 & activated(RN(s, e), v)

Indeed this property is entailed by the temporal relationships in the biological model *EGBM*.

An interpretation mapping from the cognitive model *STCM* to the biological model *STBM* for emotion reading can be defined as a specialization of the mapping φ_1 above for s = othersface(f), and extending it to the communication process:

φ_1(othersface(f)) = activated(othersface(f), 1)
φ_1(sensor_state(othersface(f))) =
activated(sensor_state(othersface(f)), 1)
φ_1(srs(othersface(f)) =
activated(SRN(othersface(f)), 1)
φ_1(imputation(othersface(f), e)) =
∃v v≥0.75 & activated(RN(othersface(f), e), v)
φ_1(preparation_state(say(your_emotion_is(e)))) =
∃v v≥0.75 & activated(PN(say(your_emotion_is(e))), v)
φ_1(effector_state(say(your_emotion_is(e)))) =
∃v v≥0.75 &
activated(effector_state(say(your_emotion_is(e))), v)

7. Relating a theory theory model

The Theory Theory perspective assumes that conclusions are drawn by reasoning from a Theory of Mind that is explicitly represented. In the cognitive model for mindreading described in [4] such an explicit representation has the form of a network of relations between (possible) cognitive states of the other person of the form leadsto(c1, c2) expressing that when a cognitive state with property c1 occurs, then a cognitive state with property c2 will occur. Such a format can be used to explicitly represent causal or temporal relations between properties of cognitive states of another agent. Reasoning methods can be identified that make use of such representations, for example reasoning methods that proceed forward in time or backward in time through the network of leadsto relations. This is the representation format adopted for the cognitive Theory Theory model *TTCM* used here. The part relevant for the context in this paper is the relationship between an emotion felt by another person and the other person's facial expression for this emotion:

leadsto(othersemotion(e), othersface(f))

That this relationship is believed is expressed by

belief(leadsto(othersemotion(e), othersface(f))

The following reasoning method can be used to derive a conclusion from the observation of a face expression:

belief(leadsto(X, Y)) & belief(Y) $\twoheadrightarrow$ belief(X)

This expresses that when a state property is believed and it is believed that it results from another state property then the other state property will be believed as well. Note that this is a temporal LEADSTO relation containing an explicitly represented leadsto relation. If this reasoning method is applied to the specific relationship here, it takes the following instantiated form (with X = othersemotion(e), Y = othersface(f)):

belief(leadsto(othersemotion(e), othersface(f))) & belief(othersface(f))) $\twoheadrightarrow$ belief(othersemotion(e))

This can be used to obtain the following reasoning pattern for emotion recognition:

I believe that the person has facial expression f
I believe that emotion e leads to facial expression f
Therefore I believe that the person has emotion e.

In this way the cognitive model from the Theory Theory perspective (*TTCM)* has been designed and specified within the LEADSTO environment. A simulation trace is shown in Figure 6; here the time delay within the temporal LEADSTO relation was taken 7 time units.

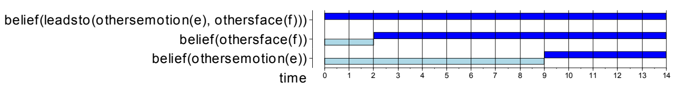

Figure 6 Simulation for emotion reading by TTCM

The following mapping relates *TTCM* and *STCM*.

φ_2(belief(othersface(f))) = srs(othersface(f))
φ_2(belief(othersemotion(e))) = imputation(othersface(f), e)
φ_2(belief(leadsto(othersemotion(e), othersface(f)) = C'

where C' is the conjunction of all relations in the cognitive model *STCM*. For X = othersemotion(e) , Y = othersface(f) it holds:

φ_2(belief(leadsto(X, Y) & belief(Y) $\twoheadrightarrow$ belief(X))
= φ_2(belief(leadsto(X, Y) & belief(Y)) $\twoheadrightarrow$ φ(belief(X))
= φ_2(belief(leadsto(X, Y)) & φ_2(belief(Y))) $\twoheadrightarrow$ φ_2(belief(X))
= C' & srs(othersface(f)) $\twoheadrightarrow$ imputation(othersface(f), e)

From this it easily can be verified that applying φ_1 after φ_2 provides an interpretation of *TTCM* in *STBM*.

8. Discussion

Models for behaviour and cognition can be designed at different levels of abstraction. Within the Agent Systems area within Artificial Intelligence, often such agent models are designed at a cognitive level. For example, the well-known BDI-model. However, it is more and more recognized that agent models need to be more 'embodied' to obtain their grounding in (physical or biological) reality. Models describing an agent's internal functioning as fully immersed in physical reality can be designed on the basis of modelling concepts that are appropriate to describe the world's dynamics, for example, concerning physical, chemical, or biological concepts and their dynamics; e.g. [8]. One option is to use such concepts to directly specify the agent model, but it is also possible to specify precisely defined relationships between the concepts used in the agent model and concepts describing biological reality..

In [6] pp. 124-132 a number of possible emotion reading models from the Simulation Theory perspective are informally sketched and discussed. The approach put forward in the current paper has drawn some inspiration from the four models sketched in [6]. Two cognitive and one biological agent models for emotion reading are considered. Based on the literature on reduction it is shown how the models can be related to each other by interpretation mappings. Such interpretation mappings show how the cognitive agent models can be grounded in a biological agent model.

9. References

[1] Baron-Cohen, S. (1995). *Mindblindness*. MIT Press.

[2] Bosse, T., Jonker, C.M., Meij, L. van der, and Treur, J., (2007). A Language and Environment for Analysis of Dynamics by Simulation. *International Journal of Artificial Intelligence Tools*, vol. 16, 2007, pp. 435-464.

[3] Bosse, T., Jonker, C.M., and Treur, J., (2008). Formalisation of Damasio's Theory of Emotion, Feeling and Core Consciousness. *Consciousness and Cognition Journal*, vol. 17, 2008, pp. 94–113.

[4] Bosse, T., Memon, Z.A., and Treur, J., (2007). Emergent Storylines Based on Autonomous Characters with Mindreading Capabilities. In: Lin, T.Y., Bradshaw, J.M., Klusch, M., Zhang, C., Broder, A., and Ho, H.(eds.), *Proc. of the Sixth Intern. Conf. on Intelligent Agent Technology, IAT'07*. IEEE Computer Society Press, 2007, pp. 207-214.

[5] Damasio, A. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Harcourt Brace, 1999.

[6] Goldman, A.I. (2006). *Simulating Minds: the Philosophy, Psychology and Neuroscience of Mindreading*. Oxford University Press.

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

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

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