

A Generic Adaptive Agent Architecture Integrating Cognitive and Affective States and their Interaction

Zulfiqar A. Memon^{1,2}, Jan Treur¹

¹VU University Amsterdam, Department of Artificial Intelligence, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

²Sukkur Institute of Business Administration (Sukkur IBA), Air Port Road Sukkur, Sindh, Pakistan

Email: {zamemon, treur}@few.vu.nl URL: <http://www.few.vu.nl/~{zamemon, treur}>

Abstract

In this paper a generic adaptive agent architecture is presented that integrates the interaction between cognitive and affective aspects of mental functioning, based on variants of notions adopted from neurological literature. It is discussed how it addresses a number of issues that have recurred in the recent literature on Cognitive Science and Philosophy of Mind.

Introduction

Recent neurological findings suggest that studying cognitive and affective aspects of mental functioning separately may not be a fruitful way to go. For example, Phelps (2006, pp. 46-47) states: ‘The mechanisms of emotion and cognition appear to be intertwined at all stages of stimulus processing and their distinction can be difficult. (...) Adding the complexity of emotion to the study of cognition can be daunting, but investigations of the neural mechanisms underlying these behaviors can help clarify the structure and mechanisms’. Similar claims have been made recently by Dolan (2002), Pessoa (2008), and others.

This paper describes a generic agent architecture that integrates cognitive and affective aspects of mental functioning and their interaction. Abstracted variants of a number of notions adopted from neurological literature served as ingredients for this architecture: cognitive, affective states, and (causal) relations between such states as having strengths expressed by real numbers, body loops and as if body loops (Damasio, 1999), the mirroring function of preparation neurons (Rizzolatti and Sinigaglia, 2008; Iacoboni, 2008; Pineda, 2009), the process of somatic marking (Damasio, 1994, 2003), and Hebbian learning (Hebb, 1949). In this paper first the generic architecture is described in some detail. Next it is shown how it addresses a number of issues that have been recurring themes in the literature on Cognitive Science and Philosophy of Mind over the last 20 years or more: subjectivity in observing and believing, simulation theory vs theory theory of mind, the reverse causation paradox in mindreading, empathic understanding, adaptivity, rationality and emotion in decision making, causal efficacy of qualia, and physical grounding.

As described in more detail in (Memon and Treur, 2009a; Memon and Treur, 2009b; Bosse, Memon, and

Treur, 2009; Memon and Treur, 2008), the proposed generic agent architecture has been applied by developing specialisations for a number of cases involving, for example, emotion generation, emotion reading, empathic understanding, believing, and trusting. The current paper describes the generic agent architecture behind these specialisations and reviews how it addresses a number of recurring themes in the literature on mind and cognition.

The Generic Adaptive Agent Architecture

A first design choice for the agent architecture was to represent (cognitive and affective) states as having certain levels or extents. This is a crucial choice as it enables dynamics of states by small changes, for example in recursive loops. A related choice is to assign strengths to causal relations between states, which also enables adaptivity. A specific agent model can be designed by using a number of such states and causal relations between them with connection strengths fixed or based on learning. In Figure 2 an overview of such basic elements is given.

Informally described theories in, for example, biological or neurological disciplines, often are formulated in terms of causal relationships or in terms of dynamical systems. To adequately formalise such a theory the hybrid dynamic modelling language LEADSTO has been developed that subsumes qualitative and quantitative causal relationships, and dynamical systems; cf. (Bosse, Jonker, Meij and Treur, 2007). Within LEADSTO the *dynamic property* or temporal relation $a \rightarrow_D b$ 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 D), state property b will occur. Below, this D will be taken as the time step Δt , and usually not be mentioned explicitly. In LEADSTO both logical and numerical calculations can be specified in an integrated manner, and a dedicated software environment is available to support specification and simulation, in which the presented agent model has been formally specified and which was used for simulation experiments.

First the process of sensing is addressed. Here W is a variable for world state properties and V for real values in the interval $[0, 1]$. For an overview, see also Figure 2.

LP1 Sensing a world state

world_state(W, V) &
connection_strength(world_state(W), sensor_state(W), ω)
→ sensor_state($W, f(\omega V)$)

LP2 Generating a sensory representation for a world state

sensor_state(W, V) &
connection_strength(sensor_state(W), srs(W), ω)
→ srs($W, f(\omega V)$)

In these specifications, the function $f(W)$ (with $W = \omega V$) can be specified by a threshold function $h(\sigma, \tau, W) = 1/(1 + \exp(-\sigma(W - \tau)))$, with steepness σ , and threshold τ , or simply by the identity function $g(W) = W$.

In principle direct connections can be made from sensory representation states to preparation states (not depicted in Figure 2), or intermediate cognitive states can be used as depicted in Figure 2. Property LP3 describes the response to a cognitive state c in the form of the preparation for a specific (bodily) reaction b . This specifies part of the recursive loop between cognitive and affective states; see Figure 1. It is calculated based on a parameterised function $f(W_1, W_2)$ of the original levels V_i (with $W_i = \omega_i V_i$).

LP3 Generating a preparation state

cognitive_state(c, V_1) & feeling(b, V_2) & preparation_state(b, V_3) & connection_strength(cognitive_state(c), preparation_state(b), ω_1) & connection_strength(feeling(b), preparation_state(b), ω_2)
→ preparation_state($b, V_3 + \gamma (f(\omega_1 V_1, \omega_2 V_2) - V_3) \Delta t$)

In different applications of the generic agent architecture, two templates of functions $f(W_1, W_2)$ have been taken:

$$g(\beta, W_1, W_2) = \beta(1 - (1 - W_1)(1 - W_2)) + (1 - \beta)W_1W_2$$
$$h(\sigma, \tau, W_1, W_2) = 1/(1 + \exp(-\sigma(W_1 + W_2 - \tau)))$$

Note that the latter formula is often used as threshold function in neural models. The first formula describes a weighted sum of two cases. The most positive case considers the two source values as strengthening each other, thereby staying under 1 : combining the imperfection rates $1 - W_1$ and $1 - W_2$ of them provides a decreased rate of imperfection $1 - (1 - W_1)(1 - W_2)$. The most negative case considers the two source values in a negative combination: combining the imperfections of them provides an increased imperfection expressed by W_1W_2 . The factor β is a characteristic that expresses the person's orientation (from 0 as most negative to 1 as most positive). The parameter γ indicates the speed of change: how flexible the state is.

A further choice was to use body loops and as if body loops for preparations of actions, adopted from (Damasio, 1999, 2003; Bosse, Jonker, and Treur, 2008). This provides a second type of recursive loop: between preparation and feeling states (see Figure 1).

Thus a combination of two loops is obtained, where connection strengths within these loops in principle are person-specific (and might be subject to learning). Depending on these personal characteristics, from a dynamic interaction within and between the two loops, for

a given stimulus an equilibrium is reached for the strength of the cognitive, preparation, and feeling state.

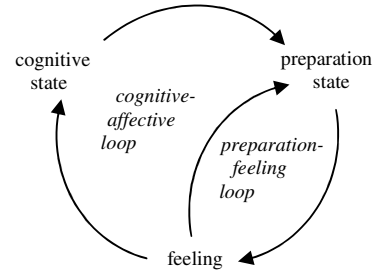


Figure 1: The two recursive loops related to a cognitive state

The existence of a connection from feeling to cognitive state can be supported by Damasio's Somatic Marker Hypothesis; cf. (Damasio, 1994, 1996, 2003; Bechara and Damasio, 2004). This is a theory on decision making which provides a central role to emotions felt. Each decision option induces (via an emotional response) a feeling which is used to mark the option. For example, when a negative somatic marker is linked to a particular option, it provides a negative feeling for that option. Similarly, a positive somatic marker provides a positive feeling for that option. Usually the Somatic Marker Hypothesis is applied to provide endorsements or valuations for options for a person's actions. However, it may be considered plausible that such a mechanism is applicable to valuations of internal cognitive states (e.g., beliefs) as well. In detail the recursive loops between preparation and feeling are specified by LP4 to LP8 (see also Figure 2). Here B is a variable for body states.

LP4 From preparation to effector state for body modification

preparation_state(B, V) &
connection_strength(preparation_state(B), effector_state(B), ω)
→ effector_state($B, f(\omega V)$)

LP5 From effector state to modified body state

effector_state(B, V) &
connection_strength(effector_state(B), body_state(B), ω)
→ body_state($B, f(\omega V)$)

LP6 Sensing a body state

body_state(B, V) &
connection_strength(body_state(B), sensor_state(B), ω)
→ sensor_state($B, f(\omega V)$)

LP7 Generating a sensory representation of a body state

sensor_state(B, V_1) & preparation_state(B, V_2) & srs(B, V_3) &
connection_strength(preparation_state(B), srs(B), ω_1) &
connection_strength(sensor_state(B), srs(B), ω_2)
→ srs($B, V_3 + \gamma (f(\omega_1 V_1, \omega_2 V_2) - V_3) \Delta t$)

LP8 From sensory representation of body state to feeling

srs(B, V) & connection_strength(srs(B), feeling(B), ω)
→ feeling($B, f(\omega V)$)

Next the property for generation of the cognitive state c is described, where both a sensory representation of w and a feeling of b play their role. This specifies another part of

the loop between cognitive and affective states, in addition to LP3 (see Figure 1). The resulting level for the cognitive state c is calculated based on a parameterised function $f(U_1, U_2)$ of the original levels V_i (with $U_i = \omega_i V_i$). Here w , c , and b are specific instances of world state, cognitive state and body state.

LP9 Generating a cognitive state

$$\begin{aligned} & \text{srs}(w, V_1) \ \& \ \text{feeling}(b, V_2) \ \& \ \text{cognitive_state}(c, V_3) \ \& \\ & \text{connection_strength}(\text{srs}(w), \text{cognitive_state}(c), \omega_1) \ \& \\ & \text{connection_strength}(\text{feeling}(b), \text{cognitive_state}(c), \omega_2) \\ & \rightarrow \text{cognitive_state}(c, V_3 + \gamma_3 (f(\omega_1 V_1, \omega_2 V_2) - V_3) \Delta t) \end{aligned}$$

Within the agent architecture the adaptive element is incorporated by making (some of the) connection strengths adaptive. From a Hebbian neurological perspective (Hebb, 1949), strengthening of connections over time may be considered plausible, when neurons involved in the connected nodes often are activated simultaneously. Therefore such a connection can be adapted based on a Hebbian learning mechanism (Hebb, 1949; Bi and Poo, 2001; Gerstner and Kistler, 2002). Based on these considerations, in the agent architecture connection strengths ω can be adapted using the following Hebbian learning rule. It takes into account a maximal connection strength l , a learning rate η , and an extinction rate ζ . A similar Hebbian learning rule can be found in (Gerstner and Kistler, 2002, p. 406). Here N_i are variables over neural states.

LP10 Hebbian learning rule template

$$\begin{aligned} & N_1(V_1) \ \& \ N_2(V_2) \ \& \ \text{connection_strength}(N_1, N_2, \omega) \ \& \\ & \text{learning_rate}(N_1, N_2, \eta) \ \& \ \text{extinction_rate}(N_1, N_2, \zeta) \\ & \rightarrow \text{connection_strength}(N_1, N_2, \omega + (\eta V_1 V_2 (1 - \omega) - \zeta \omega) \Delta t) \end{aligned}$$

By the factor $1 - \omega$ the learning rule keeps the level of ω bounded by l (which could be replaced by any number). When extinction is neglected, the upward changes during learning are proportional to both V_1 and V_2 , which in particular means that no learning takes place whenever one of them is 0, and maximal learning takes place when both are l .

How the Different Issues Are Addressed

This section addresses how the presented agent architecture addresses a number of recurring issues in the literature on cognition and mind.

- *The role of subjectivity in observing and believing*

How states and characteristics of the subject affect his or her observation and beliefs is a first theme discussed. In an idealised rational agent the generation of cognitive states might only depend on informational sources and be fully independent from non-informational aspects such as emotions. However, in real life persons may, for example, have a more optimistic or pessimistic character and affect their beliefs and other cognitive states in the sense that an optimist person strengthens beliefs that have a positive feeling associated and a pessimistic person strengthens beliefs with a negative associated feeling. Thus the strengths may depend on non-informational aspects of mental processes and related personal characteristics. The presented agent architecture allows to associate emotional responses to sensory representations that are felt (through connections LP3, LP4, LP5, LP6, LP7, LP8), and affect further cognitive processing (through LP9). This has been worked out in more detail for beliefs in (Memon and Treur, 2009a). For this specialisation in Figure 2 for the cognitive state c , the belief that w holds is taken.

- *The reverse causation paradox in mindreading*

A next issue discussed is in how far mindreading makes use of the own mental states that are counterparts of the attributed mental states. A nontrivial obstacle for the Simulation Theory perspective (e.g., Goldman, 2006) on mindreading is what can be called the *reverse causation paradox*: how to simulate a process which in principle has the reverse order compared to the causal relations used in the simulation. Further analysis reveals that his paradox in fact originates from the often made assumption that the causal relations used by the observed person flow from mental states to actions and body states, whereas the latter is what the observing agent observes. As within the observing agent this observation is the starting point for the simulation process to determine the observed person’s mental states, this would be against the direction of the causal relations used. This issue is encountered, for example, in (Goldman and Sripada, 2004; Goldman, 2006, pp. 124-132), where four possible informal emotion reading models from the Simulation Theory perspective are sketched and discussed. Partly inspired by these models, in (Memon and Treur, 2008) an emotion reading model was introduced, based on the agent architecture presented here, to address this paradox.

In (Goldman, 2006), for model 1, to resolve the reverse causation paradox, a ‘generate and test process’ for emotional states was assumed, where on the basis of a hypothesized emotional state an own facial expression is generated, and this is compared to the observed facial expression of the other person. In the assessment of this model, the unspecified hypothesis generation process for a

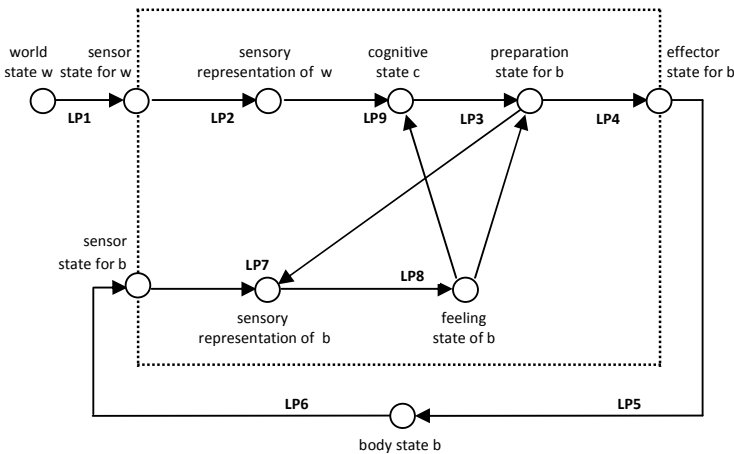


Figure 2: Basic elements of the architecture

given observed face was considered as a less satisfactory aspect. Models 2 and 3 discussed in (Goldman, 2006) are based on a notion of what he calls 'reverse simulation'. This means that for the causal relation from emotional state to (the preparation of) a facial expression which is used to generate the own facial expressions, also a reverse relation from prepared own facial expression to emotional state is assumed, which is used for the mind reading process. A point of discussion concerning these two models is that whereas the emotional states and facial expression (preparation) states used for mindreading are the same as used for the own emotions and facial expressions, the causal relations between them used in the two cases are not the same. Model 4 is based on a so-called 'mirroring process', where a correlation between the emotional state of the other person and the corresponding own emotional state is assumed, based on a certain causal chain between the two. However, the relation of such a causal chain with the causal relations used to generate the own emotional states and facial expressions is not made clear.

The model presented in (Memon and Treur, 2008) based on the agent architecture presented in the current paper, addresses the reverse causation paradox in the following manner. The picture for this specialisation is as in Figure 2, but with cognitive state *c* left out: the sensory representation of *w* has a direct connection to the preparation for *b* (which is assumed to have a mirroring function). Moreover, an additional cognitive imputation state is present to connect the own feeling and the stimulus representing the other person's face. Within this specialised model the recursive (as if) body loop addresses the problems of Goldman's model 1, as it can be viewed as an efficient and converging way of generating and testing hypotheses for the emotional states, where the (as if) body loop takes care of the generation process. Moreover, it solves the problems of models 2 and 3, as the causal chain used from facial expression to emotional state is not a reverse simulation, but just the circular causal chain (recursive body loop) which is used for generating the own responses and feeling states as well. Finally, compared to model 4, the model put forward in (Memon and Treur, 2008) can be viewed as an efficient manner to obtain a mirroring process between the emotional state of the other person on the own emotional state, based on the machinery available for the own emotional states.

- *Empathic understanding*

For humans, one of the deepest and most fundamental forms of mutual understanding is based on the notion of *empathy*; e.g., (Ickes, 1997; Preston and Waal, 2002; Decety and Jackson, 2004; Lamm, Batson, and Decety, 2007; Iacoboni, 2005, 2008). Originally (cf. Lipps, 1903) the notion of empathy was named by the German word 'einfühlung' which could be translated as 'feeling into'; e.g., (Preston and Waal, 2002). As this word indicates more explicitly, the notion has a strong relation to feeling: *empathic understanding* is a form of understanding which includes (but is not limited to) feeling what the other person feels. A particular challenge here is how to enrich

understanding of any cognitive state (such as an attention, belief, desire or intention state) of another person to a form of understanding which includes feeling the same emotion associated to this cognitive state as the other person. So a crucial aspect for empathic understanding is the way in which feelings and other mental states are interrelated. For example, a belief that something bad is to happen, may relate to feeling fear, or the belief that something good has happened may relate to feeling happiness. Another example of such a relationship is the role of cognitive elements (for example, certain thoughts) in the development, persistence and recurrence of mood disorders such as depressions; e.g., (Ingram, Miranda & Segal, 1998). So, in empathic understanding both cognitive and affective states are to be involved in their mutual relationship, of both the observed and observing person.

In (Memon and Treur, 2009b) this challenging notion of empathic understanding was addressed, based on the agent architecture presented here. The model describes how the empathic agent does not only understand another agent's cognitive state but at the same time feels the accompanying emotion. It was based on two main assumptions:

- (1) The observing agent performs mindreading using the same mental states as the observed agent
- (2) Both agents have a similar mechanism to associate feelings to a given cognitive state

Concerning assumption (1), the Simulation Theory perspective was followed; cf. (Goldman, 2006). For assumption (2) the body loop was exploited. Assuming that the observed agent and the observing agent indeed have a similar mechanism for this, makes it possible that for a given cognitive state the observing agent generates the same feeling as the observed agent. The picture for this specialisation is as in Figure 2, with *c* the cognitive state for which empathic understanding takes place. Moreover, two additional cognitive imputation states are present to connect both the cognitive state *c* and the own feeling state to the other person.

Especially in relation to assumption (2) it can be questioned to which extent the mechanisms to associate feelings to a given mental state are always the same for two persons. As it may be considered plausible that basically the mechanisms are similar, it is not difficult to imagine that both due to innate and learned individual differences in the strengths of the connections in the body loops, the extent of the empathic reaction may differ. Indeed, it is often reported that identical twins have a much higher level of mutual empathy than any two persons which are not identical twins. Moreover, it is also often considered that more empathy is shown between two persons when they have had similar experiences in life. Nevertheless, a certain extent of empathy still seems possible between persons which are not genetically identical and have not exactly the same experiences. The connection strengths may be considered parameters by which such innate and acquired individual differences can be characterised.

- *Adaptivity*

The issue of adaptivity concerns how a transparent mechanism can be obtained describing how the internal mental processes adapt to the subject's experiences with the world. Classical symbolic models usually have limited or no adaptivity, as states are often taken binary. In the agent architecture presented here such adaptivity is modelled based on the fact that states and connections are expressed by real numbers, and by use of an abstracted variant of Hebb's learning principle (Hebb, 1949; Bi and Poo, 2001; Gerstner and Kistler, 2002). It has been shown in a number of cases how this enables an agent to adapt to experiences with the world, for example, involving notions such as trust based on experiences, and learning direct emotion recognition by a form of classification of face expressions, (Bosse, Memon, and Treur, 2009; Hoogendoorn, Jaffry, and Treur, 2009; Jaffry and Treur, 2009). In (Memon, Treur, and Umair, 2009) it has been analysed how this Hebbian learning principle compares to some other well-known learning principles based on temporal discounting and memory traces.

- *Rationality and emotion in decision making*

Another recurring theme is how rational and emotional aspects in decision making can be understood in a coherent fashion. A neurological theory addressing the interaction between cognitive and affective aspects in decision making is Damasio's Somatic Marker Hypothesis; cf. (Damasio, 1994, 1996; Bechara and Damasio, 2004; Damasio, 2003). This is a theory on decision making which provides a central role to emotions felt. Within a given context, each represented decision option induces (via an emotional response) a feeling which is used to mark the option. For example, a strongly negative somatic marker linked to a particular option occurs as a strongly negative feeling for that option. Similarly, a positive somatic marker occurs as a positive feeling for that option. This theory provides an account on how emotions and rational aspects cooperate in the decision process and also explains how this can take the form of an adaptive process leading to decision making by intuition or based on 'experience'. This has been exploited in the context of the role of trust states in decision making, modelled based on the agent architecture presented here, in (Hoogendoorn, Jaffry, and Treur, 2009; Jaffry and Treur, 2009).

- *Physical grounding of agent models*

Yet another recurring theme is in how far agent models are embodied, or have some form of physical grounding. Agent models can be designed at different levels of abstraction. For example, the well-known BDI-model makes use of higher-level cognitive concepts such as beliefs, desires and intentions. 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., (Clancey, 1997; Clark, 1997). 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 artificial life like approach; e.g., (Steels and Brooks, 1995; Port and van Gelder, 1995). The agent architecture presented has adopted abstracted variants of a number of neurological principles, among which body loops, Hebbian learning, and a mirroring function of preparation states. Therefore it is easily embeddable in a model at the physiological and neurological level. Such an embedding has been described in more detail in (Memon and Treur, 2008; Treur, 2010).

- *Causal efficacy of feelings and qualia*

A recurring theme in relation to conscious experiencing or qualia is whether qualia have causal efficacy; e.g., (Kim, 1996, pp. 155-183; Duch, 2005). In neurological literature such as (Damasio 1999), qualia have been associated to certain representations of body states. Within the presented agent architecture specific types of body states and feelings associated to internal representations of them, take part in recursive loops. As such these feeling states can be said to be both causing and caused by preparation states and cognitive states. Their efficacy with respect to the preparation and cognitive states through these loops depends on the strengths of the connections in the loops. This shows that at least such feeling states have a certain extent of causal efficacy. Of course, these feeling states may be considered as being rather simplified as compared to qualia, so it is still open for discussion in how far this pattern can also count as a perspective on the question of causal efficacy of qualia.

Discussion

The generic agent architecture presented in this paper integrates the interaction between cognitive and affective aspects of mental functioning, using abstracted variants of notions adopted from neurological literature: cognitive, affective states, and (causal) relations between such states with strengths expressed by real numbers, body loops and as if body loops (Damasio, 1999), the mirroring function of preparation neurons (Rizzolatti and Sinigaglia, 2008; Iacoboni, 2008; Pineda, 2009), the process of somatic marking (Damasio, 1994, 2003), and Hebbian learning (Hebb, 1949).

This agent architecture has been applied by developing specialisations for a number of cases involving, for example, emotion generation, emotion reading, empathic understanding, believing, and trusting. The current paper describes the generic agent architecture behind these specialisations and reviews how it addresses a number of recurring themes in the literature on mind and cognition.

The architecture has been formally specified in the hybrid modelling language LEADSTO (Bosse et al., 2007). It allows for both mathematical analysis and logical

analysis (verification) as has been shown in work on specialisations of the architecture, such as (Bosse, Memon and Treur, 2009; Memon and Treur, 2009a, 2009b).

The generic agent architecture illustrates how recent developments within the neurological area can be adopted and shaped to obtain innovative design elements for agent models. It shows how cognitive modelling and artificial (general) intelligence can benefit of such developments.

References

- Bechara, A., and Damasio, A. (2004). The Somatic Marker Hypothesis: a neural theory of economic decision. *Games and Economic Behavior*, vol. 52, pp. 336-372.
- Bi, G.Q., and Poo, M.M. (2001) Synaptic Modifications by Correlated Activity: Hebb's Postulate Revisited. *Ann Rev Neurosci*, vol. 24, pp. 139-166.
- 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.
- 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.
- Bosse, T., Memon, Z.A., and Treur, J., (2009). An Adaptive Agent Model for Emotion Reading by Mirroring Body States and Hebbian Learning. In: Yang, J.-J., et al. (eds.), *Proc. of the 12th International Conference on Principles of Practice in Multi-Agent Systems, PRIMA'09*. Lecture Notes in Artificial Intelligence, vol. 5925, Springer Verlag, 2009, pp. 552-562.
- Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*. Cambridge University Press.
- Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*, Papermac, London.
- Damasio, A. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. *Philosophical Transactions of the Royal Society: Biological Sciences*, vol. 351, pp. 1413-1420
- Damasio, A. (1999). *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace, 1999.
- Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Vintage books, London, 2004.
- Decety, J. and Jackson, P.L. (2004) The functional architecture of human empathy. *Behav. Cogn. Neurosci. Rev.* 3, 71-100.
- Dolan, R.J. (2002). Emotion, Cognition, and Behavior. *Science*, vol 298, 2002, pp. 1191-1194.
- Gerstner, W., and Kistler, W.M. (2002). Mathematical formulations of Hebbian learning. *Biol. Cybern.*, vol. 87, 2002, pp. 404-415
- Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford Univ. Press.
- Goldman, A.I., and Sripada, C.S. (2004). Simulationist models of face-based emotion recognition. *Cognition*, vol. 94, pp. 193-213.
- Hebb, D.O. (1949). *The Organization of Behaviour*. John Wiley & Sons, New York, 1949.
- Hoogendoorn, M., Jaffry, S.W., and Treur, J., (2009). Modelling Trust Dynamics from a Neurological Perspective. In: *Proc. of the Second International Conference on Cognitive Neurodynamics, ICCN'09*. Springer Verlag, 2009, to appear.
- Iacoboni M. (2008). *Mirroring People: the New Science of How We Connect with Others*. New York: Farrar, Straus & Giroux
- Ickes, W. (1997). *Empathic Accuracy*. Guilford Press, New York.
- Jaffry, S.W., and Treur, J., (2009). Comparing a Cognitive and a Neural Model for Relative Trust Dynamics. In: *Proceedings of the 16th International Conference on Neural Information Processing, ICONIP'09*. Lecture Notes in Computer Science, vol. 5863. Springer Verlag, 2009, pp. 72-83
- Kim, J. (1996). *Philosophy of Mind*. Westview Press.
- Lamm, C., Batson, C.D., and Decety, J. (2007). The neural basis of human empathy – effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosci.*, vol. 19, 2007, pp. 42-58.
- Lipps, T. (1903) Einfühlung, innere Nachahmung und Organempfindung. *Archiv für die gesamte Psychologie*, vol. 1, pp. 465-519.
- Memon, Z.A., and Treur, J., (2008). Cognitive and Biological Agent Models for Emotion Reading. In: Jain, L., et al. (eds.), *Proceedings of the 8th IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT'08*. IEEE Computer Society Press, 2008, pp. 308-313.
- Memon, Z.A., and Treur, J., (2009a). Modelling the Reciprocal Interaction between Believing and Feeling from a Neurological Perspective. In: N. Zhong et al. (eds.), *Proc. of the First Intern. Conf. on Brain Informatics, BI'09*. Lecture Notes in Artificial Intelligence, vol. 5819. Springer Verlag, 2009, pp. 13-24.
- Memon, Z.A., and Treur, J., (2009b). Designing Social Agents with Empathic Understanding. In: Nguyen, N.T., et al. (eds.), *Proc. of the First International Conference on Computational Collective Intelligence, ICCCI'09*. Lecture Notes in Artificial Intelligence, vol. 5796. Springer Verlag, 2009, pp. 279-293.
- Memon, Z.A., Treur, J., and Umair, M., (2009). A Comparative Analysis on Adaptive Modelling of Induced Feelings. In: *Proc. of the Second International Conference on Cognitive Neurodynamics, ICCN'09*. Springer Verlag, 2009, to appear.
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews: Neuroscience*, vol. 9, 2008, pp. 148-158.
- Pelphs, E.A. (2006). Emotion And Cognition: Insights from Studies of the Human Amygdala. *Annu. Rev. Psychol.* 2006. 57:27-53
- Pineda, J.A. (ed.), (2009). *Mirror Neuron Systems: the Role of Mirroring Processes in Social Cognition*. Humana Press Inc.
- Port, R.F., Gelder, T. van (eds.), (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. MIT Press, Cambridge, Mass, 1995.
- Preston, S.D. and Waal, F.B.M. de (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1-72.
- Rizzolatti, G, and Sinigaglia, C., (2008). *Mirrors in the Brain: How Our Minds Share Actions and Emotions*. Oxford University Press, 2008.
- Steels, L. & Brooks, R. (1995). *The artificial life route to artificial intelligence: Building embodied, situated agents*. Erlbaum.
- Duch, W. (2005), Brain-inspired conscious computing architecture. *Journal of Mind and Behavior* 26(1-2) (2005) 1-22
- Treur, J., (2010). On the Use of Reduction Relations to Relate Different Types of Agent Models. *Web Intelligence and Agent Systems Journal*, 2010, to appear.