

A Cognitive Agent Model Incorporating Prior and Retrospective Ownership States for Actions

Jan Treur

VU University Amsterdam, Agent Systems Research Group
De Boelelaan 1081, 1081 HV, Amsterdam, The Netherlands
treur@cs.vu.nl <http://www.cs.vu.nl/~treur>

Abstract

The cognitive agent model presented in this paper generates prior and retrospective ownership states for an action based on principles from recent neurological theories. A prior ownership state is affected by prediction of the effects of a prepared action, and exerts control by strengthening or suppressing actual execution of the action. A retrospective ownership state depends on whether the sensed consequences co-occur with the predicted consequences, and is the basis for acknowledging authorship of actions, for example, in social context. It is shown how poor action effect prediction capabilities can lead to reduced retrospective ownership states, as in persons suffering from schizophrenia.

1 Introduction

In the neurological literature the notion of ownership of an action has received much attention: in how far does a person attribute an action to him or herself, or to another person. For example, persons suffering from schizophrenia may easily attribute self-generated actions to (real or imaginary) other persons. One of the issues that plays an important role both in the execution decisions for an action, and in its attribution, is the prediction of the (expected) effects of the action, based on internal simulation starting from the preparation of the action (e.g., [Wolpert, 1997; Haggard, 2008]). If these predicted effects are satisfactory, this may entail a ‘go’ decision for the execution of the action, thus exerting control over action execution. In contrast, less satisfactory predicted effects may lead to vetoing a prepared action: a ‘no go’ decision.

Predicted action effects also play an important role in attribution of the action to an agent after it has been performed. In neurological research it has been found that poor predictive capabilities are a basis for false attributions of actions, for example, for patients suffering from schizophrenia; (e.g., [Synofzik *et al.*, 2010; Voss *et al.*, 2010]). The idea developed over the years is that co-

occurrence of predicted effects and sensed actual effects (after execution of the action) is an important condition for proper retrospective self-attribution of a self-generated action (e.g., [Feinberg, 1978; Frith, 1992; Wolpert, 1997; Frith *et al.*, 2000; Moore and Haggard, 2008]). A peculiar aspect here is that within the process the predicted effect suppresses the sensed actual effect (e.g., [Blakemore *et al.*, 1999; Blakemore *et al.*, 2000; Fournier *et al.*, 2002]). In recent years it has been put forward that the predicted effect and the sensed actual effect are not simply compared or matched, as claimed in the so-called ‘comparator model’ in earlier literature such as [Wolpert, 1997; Frith, 1992; Frith *et al.*, 2000], but in fact are added to each other in some integration process (e.g., [Moore and Haggard, 2008; Synofzik *et al.*, 2010; Voss *et al.*, 2010]).

The cognitive agent model presented in this paper is based on the perspective put forward in the latter recent literature. In designing this model, in line with this literature a useful distinction was made between prior ownership states, among others based on prediction of effects of a prepared action, and retrospective ownership states, for which in addition the monitored execution of the action and the sensed actual effects play an important role. Prior ownership states play an important role in controlling the actual execution of actions (go/no-go decisions, vetoing), whereas retrospective ownership states are important for acknowledging authorship of an action in a social context, but also may play a role in reflection on one’s own functioning and personal learning and development (e.g., learning from less optimal choices).

In this paper, in Section 2 the cognitive agent model is presented, based on neurological principles as mentioned above. Section 3 illustrates the model by presenting four different scenarios: action execution with proper self-attribution, vetoing an action because of unsatisfactory predicted effects, execution of an action with false attribution due to poor predictive capabilities as happens in schizophrenia, and mirroring an observed action of another agent and properly attributing it to the other agent. Finally, Section 4 is a discussion.

2 The Cognitive Agent Model

The issues and perspectives briefly reviewed in the introduction have been used as a basis for the neurologically inspired cognitive agent model presented below (for an overview, see Figure 1), more specifically the following have been incorporated:

- (1) action effect *prediction* from preparation of an action a to sensory representation of effect b
- (2) *suppressing* the sensory representation of effect b after action a is initiated
- (3) a *prior ownership* state depends on preparation for the action, predicted effects, and context
- (4) a *retrospective ownership* state depends on co-occurrence of predicted action effects and action effects sensed afterwards
- (5) a prior ownership state exerts *control over the execution* of a prepared action (go/no-go decision)
- (6) a retrospective ownership state is the basis for *acknowledging authorship* of the action, for example, in social context

In the model s denotes a stimulus, c a context, a an action, and b a world state affected by the action. Examples of contexts are another agent B which is observed, or the agent self. The effect state b is considered to be positive for the agent (e.g., in accordance with a goal).

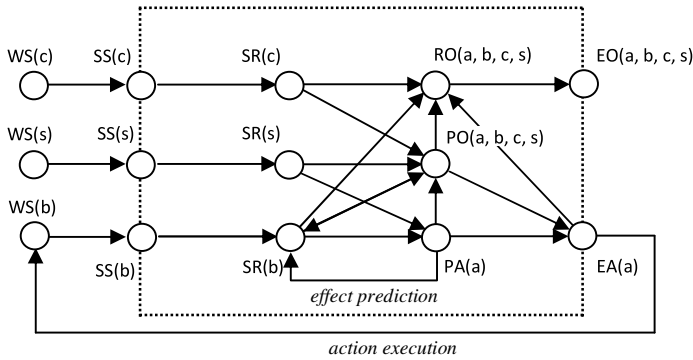


Figure 1: Overview of the cognitive agent model

The state properties used in the model are summarised in Table 1. As expressed in (3) and (4) above, the cognitive agent model distinguishes prior and retrospective ownership states for actions, indicated by $PO(a, b, c, s)$ and $RO(a, b, c, s)$, respectively (see Figure 1). These states are taken specific for a given action a , effect b , context c , and stimulus s (triggering preparation of a). When the context c is self, an ownership state for c indicates self-ownership attribution, whereas for context c an observed agent B, it indicates ownership attributed to B. Note that the stimulus s triggering preparation of action a can be of any type; for social scenarios, it can be taken as a body state (e.g., face expression) of the other agent B. An action effect state b can be any state of the world (possibly including body states).

In accordance with (3) above, the prior ownership state $PO(a, b, c, s)$ is affected by the preparation state $PA(a)$ for the action a , the sensory representation $SR(b)$ of the (predicted) effect b , the sensory representation $SR(s)$ of the stimulus s ,

and the sensory representation $SR(c)$ of the context c ; see the four arrows to $PO(a, b, c, s)$ in Figure 1. Similarly, as expressed in (4) above, the retrospective ownership state $RO(a, b, c, s)$ is affected by the sensory representation $SR(c)$ of the context c , the sensory representation $SR(b)$ of the effect b of the action, the prior ownership state $PO(a, b, c, s)$, and the execution $EA(a)$ of the action a ; see the arrows to $RO(a, b, c, s)$ in Figure 1.

Action prediction, expressed in (1) above, is modelled by the connection from the action preparation $PA(a)$ to the sensory representation $SR(b)$ of the effect b . Suppression of the sensory representation of the effect, expressed as (2) above, is modelled by the (inhibiting) connection from the prior ownership state $PO(a, b, c, s)$ to sensory representation $SR(b)$. The control exerted by the prior ownership state, expressed in (5) above, is modelled by the connection from $PO(a, b, c, s)$ to $EA(a)$. Finally, acknowledging of ownership, expressed in (6) above, is modelled by the connection from the retrospective ownership state $RO(a, b, c, s)$ to the communication effector state $EO(a, b, c, s)$.

| notation | description |
|----------------|---|
| WS(W) | world state W (W is a context c, stimulus s, or effect b) |
| SS(W) | sensor state for W |
| SR(W) | sensory representation of W |
| PA(a) | preparation for action a |
| EA(a) | execution of action a |
| PO(a, b, c, s) | prior ownership state for action a with b, c, and s |
| RO(a, b, c, s) | retrospective ownership state for a with b, c, and s |
| EO(a, b, c, s) | communication of ownership of a with b, c, and s |

Table 1: State properties used

Connections between state properties (the arrows in Figure 1) have weights ω_k , as indicated in Table 2. In this table the column LP refers to the (temporally) Local Properties LP1 to LP9 presented below. A weight ω_k has a value between -1 and 1 and may depend on the specific context c , stimulus s , action a and/or effect state b involved. By varying these connection strengths, different possibilities for the repertoire offered by the model can be realised. Note that usually weights are assumed non-negative, except for the inhibiting connections, such as ω_{20} , which models suppression of the sensory representation of effect b .

| from states | to state | weights | LP |
|-------------------------------------|----------------|--|-----|
| SS(W) | SR(W) | ω_1 | LP1 |
| PA(a), PO(a, b, self, s), SS(b) | SR(b) | $\omega_2, \omega_{20}, \omega_3$ | LP2 |
| SR(s), SR(b) | PA(a) | ω_4, ω_5 | LP3 |
| SR(c), SR(s), SR(b), PA(a) | PO(a, b, c, s) | $\omega_6, \omega_7, \omega_8, \omega_9$ | LP4 |
| PO(a, b, self, s), PA(a) | EA(a) | ω_{10}, ω_{11} | LP5 |
| EA(a) | WS(b) | ω_{12} | LP6 |
| WS(W) | SS(W) | ω_{13} | LP7 |
| SR(c), SR(b), PO(a, b, c, s), EA(a) | RO(a, b, c, s) | $\omega_{14}, \omega_{15}, \omega_{16}, \omega_{17}$ | LP8 |
| AO(a, b, c, s) | EO(a, b, c, s) | ω_{18} | LP9 |

Table 2: Overview of the connections and their weights

Below, the dynamics following the connections between the states in Figure 1 are described in more detail. This is done for each state by a dynamic property specifying how the activation value for this state is updated based on the activation values of the states connected to it (the incoming arrows in Figure 1).

The cognitive agent model has been computationally formalised in this way using the hybrid modeling language LEADSTO; cf. [Bosse *et al.*, 2007]. Within LEADSTO a dynamic property or temporal causal relation $a \rightarrow b$ denotes that when a state property a (or conjunction thereof) occurs, then after a certain time delay, state property b will occur. Below, this delay will be taken as a uniform time step Δt . Each time first a semiformal description is given, and next a formal specification in the hybrid LEADSTO format. Parameter γ is a speed factor, indicating the speed by which an activation level is updated upon received input from other states.

During processing, each state property has a strength represented by a real number between 0 and 1; variables V (possibly with subscripts) run over these values. In dynamic property specifications, this is added as a last argument to the state property expressions (an alternative notation activation(p , V) with p a state property has not been used for the sake of notational simplicity).

Below, f is a function for which different choices can be made, for example, the identity function $f(W) = W$ or a combination function based on a continuous logistic threshold function of the form

$$th(\sigma, \tau, X) = \left(\frac{1}{1 + e^{-\sigma(X - \tau)}} - \frac{1}{1 + e^{\sigma\tau}} \right) (1 + e^{-\sigma\tau})$$

with σ a steepness and τ a threshold value. Note that for higher values of $\sigma\tau$ (e.g., σ higher than $20/\tau$) this threshold function can be approximated by the simpler expression:

$$th(\sigma, \tau, X) = \frac{1}{1 + e^{-\sigma(X - \tau)}}$$

In the example simulations, for the states that are affected by only one state (i.e., in LP1, LP6, LP7, LP9), f is taken the identity function $f(W) = W$, and for the other states f is a combination function based on the logistic threshold function: $f(X_1, X_2) = th(\sigma, \tau, X_1 + X_2)$, and similarly for more arguments. In this choice common practice is followed, but other types of combination functions might be used as well.

The first property LP1 describes how sensory representations are generated for context c and stimulus s , and effect state b (together indicated by variable W).

LP1 Sensory representation for a sensor state

If the sensor state for W has level V_1
and the sensory representation of W has level V_2
then after duration Δt the sensory representation of W will have level $V_2 + \gamma [f(\omega_1 V_1) - V_2] \Delta t$.
SS(W , V_1) & SR(W , V_2) \rightarrow SR(W , $V_2 + \gamma [f(\omega_1 V_1) - V_2] \Delta t$)

The sensory representation of an effect state b as described by property LP2 is not only affected by a corresponding sensor state for b (which in turn is affected by the world state), as in LP1, but also by two action-related states:

- via the predictive loop by a preparation state, to predict the effect b of a prepared action a (see (1) above)
- by an inhibiting connection from the prior self-ownership state, to suppress the sensory representation of the effect b of the action a , once it is initiated (see (2) above)

This is expressed in dynamic property LP2. Note that for this suppressing effect the connection weight ω_{20} from prior ownership state for action a to sensory representation for effect b is taken negative, for example $\omega_{20} = -1$.

LP2 Sensory representation for an effect state

If the preparation state for action a has level V_1
and the prior self-ownership of action a for b , self, and s has level V_2
and the sensor state for state b has level V_3
and the sensory representation of state b has level V_4
then after duration Δt the sensory representation of state b will have level $V_4 + \gamma [f(\omega_2 V_1, \omega_{20} V_2, \omega_3 V_3) - V_4] \Delta t$.
PA(a , V_1) & PO(a , b , self, s , V_2) & SS(b , V_3) & SR(b , V_4)
 \rightarrow SR(b , $V_4 + \gamma [f(\omega_2 V_1, \omega_{20} V_2, \omega_3 V_3) - V_4] \Delta t$)

Preparation for action a is affected by a sensory representation of stimulus s (triggering the action), and also strengthened by predicted effect b of the action:

LP3 Preparing for an action

If sensory representation of s has level V_1
and sensory representation of b has level V_2
and the preparation for action a has level V_3
then after duration Δt the preparation state for action a will have level $V_3 + \gamma [f(\omega_4 V_1, \omega_5 V_2) - V_3] \Delta t$.
SR(s , V_1) & SR(b , V_2) & PA(a , V_3)
 \rightarrow PA(a , $V_3 + \gamma [f(\omega_4 V_1, \omega_5 V_2) - V_3] \Delta t$)

Prior ownership of an action a is generated by LP4 (see (3) above).

LP4 Generating a prior ownership state

If the sensory representation of context c has level V_1
and the sensory representation of s has level V_2
and sensory representation of b has level V_3
and the preparation for action a has level V_4
and prior ownership of a for b , c , and s has level V_5
then after duration Δt prior ownership of a for c , s , and b will have level $V_5 + \gamma [f(\omega_6 V_1, \omega_7 V_2, \omega_8 V_3, \omega_9 V_4) - V_5] \Delta t$.
SR(c , V_1) & SR(s , V_2) & SR(b , V_3) & PA(a , V_4) & PO(a , b , c , s , V_5)
 \rightarrow PO(a , b , c , s , $V_5 + \gamma [f(\omega_6 V_1, \omega_7 V_2, \omega_8 V_3, \omega_9 V_4) - V_5] \Delta t$)

In case the context c is self, the prior ownership state strengthens the initiative to perform a as a self-generated action: executing a prepared action depends on whether a prior self-ownership state (for the agent self) is available for this action (see (5) above). This models control over the actual execution of the action (go/no-go decision) and can, for example, be used to veto the action in a late stage of preparation. This is modelled by LP5.

LP5 Action execution

If prior ownership of a for b, self, and s has level V_1
and preparation for action a has level V_2
and the action execution state for a has level V_3
then after duration Δt the action execution state for a will have
level $V_3 + \gamma [f(\omega_{10}V_1, \omega_{11}V_2) - V_3] \Delta t$.
 $PO(a, b, self, s, V_1) \ \& \ PA(a, V_2) \ \& \ EA(a, V_3)$
 $\rightarrow EA(a, V_3 + \gamma [f(\omega_{10}V_1, \omega_{11}V_2) - V_3] \Delta t)$

Property LP6 describes in a straightforward manner how execution of action a affects the world state b.

LP6 From action execution to effect state

If the execution state for action a has level V_1 ,
and world state b has level V_2
then after Δt world state b will have
level $V_2 + \gamma [f(\omega_{12}V_1) - V_2] \Delta t$.
 $EA(a, V_1) \ \& \ WS(b, V_2) \rightarrow WS(b, V_2 + \gamma [f(\omega_{12}V_1) - V_2] \Delta t)$

The following property models how sensor states are updated. It applies to stimulus s, effect b, and context c (indicated by variable W).

LP7 Generating a sensor state for a world state

If world state W has level V_1
and the sensor state for W has level V_2
then after Δt the sensor state for W will have
level $V_2 + \gamma [f(\omega_{13}V_1) - V_2] \Delta t$.
 $WS(W, V_1) \ \& \ SS(W, V_2) \rightarrow SS(W, V_2 + \gamma [f(\omega_{13}V_1) - V_2] \Delta t)$

A retrospective ownership state takes into account the prior ownership, the execution of the action, the context, and the sensory representation of the action's effect (see (4) above):

LP8 Generating a retrospective ownership state

If the sensory representation of context c has level V_1 ,
and the sensory representation of effect state b has level V_2
and prior ownership of a for b, c, and s has level V_3
and the execution state for action a has level V_4
and retrospective ownership of a for b, c, and s has level V_5
then after Δt retrospective ownership of a for b, c, and s will have
level $V_5 + \gamma [f(\omega_{14}V_1, \omega_{15}V_2, \omega_{16}V_3, \omega_{17}V_4) - V_5] \Delta t$.
 $SR(c, V_1) \ \& \ SR(b, V_2) \ \& \ PO(a, b, c, s, V_3) \ \& \ EA(a, V_4) \ \& \ RO(a, b, c, s, V_5)$
 $\rightarrow RO(a, b, c, s, V_5 + \gamma [f(\omega_{14}V_1, \omega_{15}V_2, \omega_{16}V_3, \omega_{17}V_4) - V_5] \Delta t)$

Note that LP8 applies for context self as context, but also to an observed other agent B. For an observed other agent as context the connection strength ω_{17} in LP8 is assumed 0 or negative; in the simulated scenarios discussed in Section 3 it was taken $\omega_{17} = -1$. The communication to attribute authorship (to any context c) depends on the retrospective ownership state as specified in LP9 (see (6) above).

LP9 Communication of ownership awareness

If retrospective ownership of a for b, c, and s has level V_1 ,
and communication of a for b, c, and s has level V_2
then after duration Δt communication of a for b, c, and s will
have level $V_2 + \gamma [f(\omega_{18}V_1) - V_2] \Delta t$.
 $RO(a, b, c, s, V_1) \ \& \ EO(a, b, c, s, V_2)$
 $\rightarrow EO(a, b, c, s, V_2 + \gamma [f(\omega_{18}V_1) - V_2] \Delta t)$

3 Simulation of Example Scenarios

In this section simulations are discussed for a number of example scenarios, which all involve the occurrence of a preparation state for an action a, triggered by some stimulus s. These scenarios relate to phenomena in the literature, as discussed in Section 1. They have been generated using numerical software. First a scenario is addressed where the prepared action has *satisfactory predicted effects* and therefore is executed; in this case both prior and retrospective self-ownership states occur. Next, a case is considered where the prepared action lacks positive predicted effects, and is therefore not executed: a *no-go decision*, or *vetoing*. Only a rather low prior self-ownership state is developed and no retrospective self-ownership state. In the third case, a *poor action prediction capability* is modelled, which leads to a not very high prior self-ownership state, but sufficient to actually execute the prepared action. In this case no retrospective self-ownership state occurs, as the sensory representation of the effect stays low. In the fourth case, the stimulus triggering the action preparation is the observation of another agent performing the action. In this case a low prior *self-ownership* state is generated, but high prior and retrospective *other-ownership* states. This models *mirroring* of and attribution to the other agent.

3.1 Normal Execution and Attribution of an Action

The first case considered describes a situation where the context c is the agent itself, and a stimulus s occurs. The action effect b is considered positive for the agent. The scenario is as follows:

Scenario 1

- external stimulus s occurs and triggers preparation of action a
- based on the preparation state for a the sensory representation of predicted effect b of a is generated
- based on this positive predicted effect and the other states a prior self-ownership state for action a is generated
- this prior self-ownership state for action a leads to actual execution of action a
- the execution of a affects b in a positive manner and, via sensing, also the sensory representation of b
- at the same time the sensory representation of b is suppressed due to the prior self-ownership state
- based on the generated states, after the execution of action a the agent develops a retrospective self-ownership state
- finally the agent communicates this self-ownership

The simulation of this scenario is shown in Figure 2. Parameter values used (in all scenarios set by hand) can be found in Table 3. The step size taken is $\Delta t = 0.25$. All relevant connection strengths not mentioned in this table were taken 1. The slow value 0.3 for γ was applied for external processes (action execution, effect generation and effect sensing) modelled by LP5, LP6, and LP7, and the fast value 0.6 for γ for the internal processes modelled by the other LP's.

In Figure 2 it is shown that (after sensing the stimulus), the preparation for action *a* starts around time point 2, and the representation of the predicted effect *b* around time point 3. As a result of this, around time point 6 the prior self-ownership state starts to develop, which leads to the execution of the action, starting around time point 7. In the meantime the representation of the action effect *b* is suppressed (cf. [Blakemore *et al.*, 1999; Blakemore *et al.*, 2000; Fourneret *et al.*, 2002]), causing a dip in the graph around time point 10. When the execution of the action *a* is taking place, the sensing of its effect *b* in the world has a positive impact on the representation of *b* from time point 10 on, and the retrospective self-ownership state is developed, starting from around time 15. After this, the communication of the self-ownership takes place from time point 20.

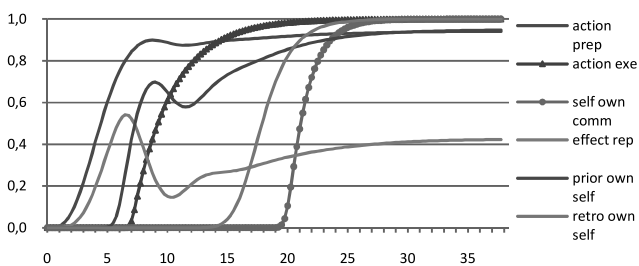


Figure 2: Executing an action with ownership states (scenario 1)

Note that in this case both the prior and the retrospective self-ownership state reach levels close to 1 (prior self-ownership approaching 0.95, and retrospective self-ownership approaching 1). Moreover, note that when the stimulus is taken away, all activation levels will go down to 0, and will come up again when the stimulus reoccurs.

| connections | | threshold and steepness for state | τ | σ |
|---------------|-----------|-----------------------------------|--------|----------|
| ω_3 | 0.5 | action preparation | 0.4 | 4 |
| ω_{20} | -1 | effect representation | 0.2 | 4 |
| ω_2 | 0.8 | action execution | 1.2 | 20 |
| ω_4 | 0.8 | prior self-ownership | 3 | 8 |
| ω_5 | 0.8 | retrospective self-ownership | 3 | 20 |
| γ | 0.6 / 0.3 | self-ownership communication | 0.8 | 40 |

Table 3: Parameter values for the first scenario

3.2 Vetoing a Prepared Action

The second case considered describes a situation similar to the previous one (the context *c* is the agent itself, and a stimulus *s* occurs), but where the action *a* triggered by stimulus *s* has an effect *b'* which is not particularly positive for the agent; here *a* hardly has an impact on effect *b* which would have been positive. Prediction capabilities are assumed correct in this case, so no high level of *b* is correctly predicted for *a*. For this situation the following variation on the previous scenario is considered:

Scenario 2

- external stimulus *s* occurs and triggers preparation of action *a*
- based on the preparation state of *a* only a low level for the sensory representation of predicted effect *b* of *a* is generated
- based on this low predicted effect *b* and the other states a low level of a prior self-ownership state for action *a* is generated
- the low prior self-ownership state for *a* does not lead to actual execution of action *a*; the action *a* can be considered vetoed
- the agent develops no retrospective self-ownership state for *a*
- the agent does not communicate self-ownership for *a*

The simulation of this scenario is shown in Figure 3. This scenario was modelled by taking the connection strength for the prediction of effect *b* for action *a* low: $\omega_2 = 0.3$ instead of 0.8. Values for the other parameters were the same as in Table 3.

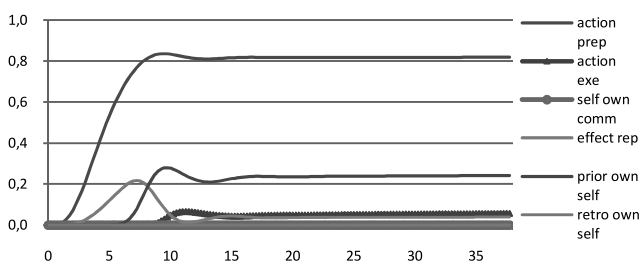


Figure 3: Vetoing an action with no positive prediction (scenario 2)

In Figure 3 it is shown that (after sensing the stimulus), again the preparation for action *a* starts around time point 2, and the representation of the predicted effect *b* around time point 3. However, the predicted effect is much lower compared to the previous scenario. As a result of this low prediction, the prior self-ownership state starting to develop around time point 6, also stays at a low level. Therefore the execution of the action also stays very low. Due to these circumstances, no retrospective self-ownership state and no communication of self-ownership occur.

3.3 Effects of Poor Prediction; Schizophrenia Case

The third case considered describes a situation where again the context *c* is the agent itself, and stimulus *s* occurs. The action effect for action *a* is *b*, which in principle is positive for the agent, like in the first situation above. However, due to poor prediction capabilities this effect is not (fully) internally predicted. This is what is assumed to happen in patients with schizophrenia, as discussed, for example, in [Synofzik *et al.*, 2010; Voss *et al.*, 2010]. For this situation the following scenario is considered:

Scenario 3

- stimulus *s* occurs and triggers preparation of action *a*
- based on the preparation state for *a* only a relatively low level of the sensory representation of the predicted effect *b* of *a* is generated, due to poor prediction capabilities
- based on this relatively low predicted effect and the other states a relatively low level of a prior self-ownership state for action *a* is generated

- this prior self-ownership state level for action a is still sufficient to lead to actual execution of action a
- the execution of a affects b in a positive manner and (via sensing) the sensory representation of b
- the sensory representation of b is suppressed to a certain extent due to the (relatively low) prior self-ownership state
- due to the relatively low level for the sensory representation of effect b (and prior self-ownership state) the agent develops no retrospective self-ownership state for action a
- the agent does not communicate self-ownership for action a

The simulation of this scenario is shown in Figure 4. This scenario was modelled by taking the connection strength for the prediction of effect b for action a moderately low: $\omega_2 = 0.4$. Values for the other parameters were again the same as in Table 3. For this case $\Delta t = 0.1$ was taken instead of 0.25, and the simulation was shown up to time point 75.

In Figure 4 it is shown that as in the previous scenarios, the preparation for action a starts around time point 2, and the representation of the predicted effect b around time point 3. The predicted effect is substantially lower compared to the first scenario, but higher than in the second scenario. As a result of this moderately low prediction, the prior self-ownership state, starting to develop around time point 6, also stays at a moderate level (first around 0.4, later going up to almost 0.7); this is substantially higher than in the second scenario where the lower level led to a veto for the action. Therefore, in contrast to the previous scenario, this level turns out high enough for the execution of the action starting around time point 9. Nevertheless, only a low level of the retrospective self-ownership state is developed (becoming approximately 0.15), and no communication of self-ownership takes place.

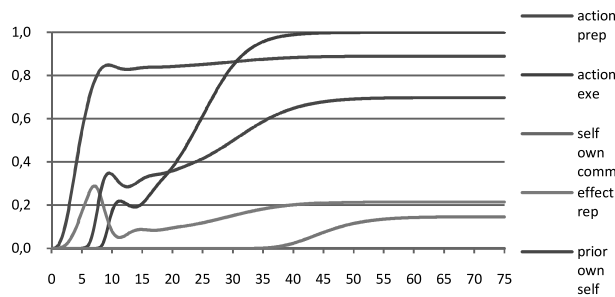


Figure 4: Poor prediction implies no self-ownership (scenario 3)

3.4 Mirroring Another Agent

The fourth case describes a situation where the context c is *another agent*, and the stimulus s is the observation of the other agent performing action a. The action effect for action a is b, and is predicted in a correct manner, as in the first scenario. The scenario for this fourth case is as follows:

Scenario 4

- external stimulus s which is an observed action a performed by another agent triggers for preparing action a (mirroring)

- based on the preparation state for a the sensory representation of the predicted effect b of a is generated
- based on this predicted effect and the other states (among which the other agent as context) a high level of a prior *other*-ownership state for action a is generated, and a low level of a prior *self*-ownership state
- the low prior self-ownership state for action a leads to no actual execution of action a (vetoing)
- as the prior self-ownership state has a low level, not much suppression of the representation of effect b takes place
- based on the generated states, the agent develops a retrospective other-ownership state, and no retrospective self-ownership state
- finally the agent communicates this other-ownership

The simulation of this scenario is shown in Figure 5. This scenario was modelled by taking the connection strength ω_4 for mirroring from the specific stimulus representation (observed action) to preparation state 0.5. The connection strength ω_2 for the prediction of effect b for action a is 0.8, as in the first scenario. The threshold and steepness values for prior and retrospective other-ownership states were taken 3 and 8, resp. 2.4 and 20. For this case $\Delta t = 0.25$ was taken. Values for the other parameters were the same as for the first scenario.

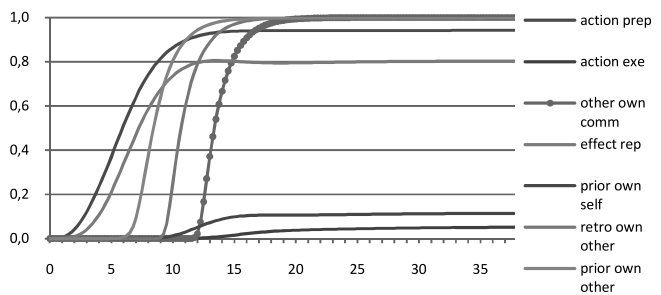


Figure 5: Mirroring another agent (scenario 4)

In Figure 5 it is shown that after sensing the observed action, as in the first scenario the preparation for action a starts around time point 2, and the representation of the predicted effect b around time point 3. As a result of this, around time point 6 the prior *other*-ownership state starts to develop, whereas the prior *self*-ownership state stays very low. Therefore execution of a is suppressed. After time point 9 also the retrospective other-ownership state is generated, which leads to communication of other-ownership after time point 12. All other states stay low.

4 Discussion

The cognitive agent model presented in this paper incorporates mechanisms for prior and retrospective ownership states, based on principles from recent neurological theories, in particular from [Moore and Haggard, 2008; Synofzik *et al.*, 2010; Voss *et al.*, 2010]. In the model a prior ownership state is affected by prediction

of the effects of the action. Actual execution of the action and sensing of its effects can lead to a retrospective ownership state, in particular, when the sensed effects co-occur with the predicted effects. As a prior ownership state may lead to actual execution of the action, it plays an important role as control of the execution of prepared actions. A retrospective ownership state is the basis for acknowledging authorship of an action, for example, in social context, or in a reflection context.

In simulated scenarios it was shown how a number of known phenomena can occur. For example, scenarios were shown for vetoing a prepared action due to unsatisfactory predicted effects, and for mirroring an observed action performed by another agent, without imitating the action. Moreover, it was shown how poor action effect prediction capabilities can lead to reduced retrospective ownership states (as, for example, is shown in persons suffering from schizophrenia), and may easily lead to attribution of the self-generated action to another real or imaginary person.

The model distinguishes itself from existing approaches such as in [Wolpert, 1997; Frith, 1992; Frith *et al.*, 2000], among others in that (1) instead of comparison of predicted and sensed effects, the predicted and sensed effects are integrated and provide a kind of combined level, as also indicated in, for example [Moore and Haggard, 2008; Synofzik *et al.*, 2010; Voss *et al.*, 2010], (2) following [Moore and Haggard, 2008] a distinction was made between prior and retrospective ownership states, and (3) both self-ownership and other-ownership are covered. These are also differences with [Hindriks *et al.*, 2011], which does not take the neurological angle as a point of departure, as in the current paper.

The obtained cognitive model can be used as a basis for the design of human-like virtual agents in simulation-based training or in gaming or virtual stories. For the first type of application the idea is to develop a virtual patient based on the model so that, for example, a psychiatrist or psychotherapist (e.g., during his or her education) can gain insight in the processes in certain types of patients, or it can be used by a therapist to analyse how a certain form of therapy can have its effect on these processes. For the second type of application the idea is to design a system for agent-based virtual stories in which, for example, persons with deviations in ownership states play a role (e.g., persons suffering from schizophrenia, and due to that attribute their own actions to other real or imaginary persons), which can be based on the presented model. In [Treur and Umair, 2011] it is shown how so-called inverse mirroring enables a person to attribute an action to an imaginary person.

References

- [Blakemore *et al.*, 1999] Blakemore, S.-J., Frith, C.D., and Wolpert, D.M., Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *J. of Cognitive Neuroscience*, 11: 551–559, 1999.
- [Blakemore *et al.*, 2000] Blakemore, S.-J., Wolpert, D.M., and Frith, C.D., Why can't you tickle yourself? *Neuroreport*, 11: 11-16, 2000.
- [Bosse *et al.*, 2007] Bosse, T., Jonker, C.M., Meij, L. van der, and Treur, J., A Language and Environment for Analysis of Dynamics by Simulation. *Intern. J. of Artificial Intelligence Tools*, 16: 435-464, 2007.
- [Feinberg, 1978] Feinberg, I., Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, 4: 636–640, 1978.
- [Fournieret *et al.*, 2002] Fournieret, P., de Vignemont, F., Franck, N., Slachevsky, A., Dubois, B., and Jeannerod, M., Perception of self-generated action in schizophrenia. *Cogn. Neuropsychiatry*, 7: 139 -156, 2002.
- [Frith, 1992] Frith, C.D., *The cognitive neuro-psychology of schizophrenia*. Hove, UK: Lawrence Erlbaum, 1992.
- [Frith *et al.*, 2000] Frith, C.D., Blakemore, S., and Wolpert, D., Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, 31: 357–363, 2000.
- [Haggard, 2008] Haggard, P., Human volition: towards a neuroscience of will. *Nature Neuroscience Reviews*, 8: 934-946, 2008.
- [Hindriks *et al.*, 2011] Hindriks, K., Wiggers, P., Jonker, C.M. and Haselager, W., Towards A Computational Model of the Self-Attribution of Agency. In: *Proc. of the 24th Intern. Conf. on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE'11*. Lecture Notes in AI, Springer Verlag, 2011.
- [Moore and Haggard, 2008] Moore, J., and Haggard, P., Awareness of action: Inference and prediction. *Consciousness and Cognition*, 17: 136–144, 2008
- [Synofzik *et al.*, 2010] Synofzik, M., Thier, P., Leube, D.T., Schlotterbeck, P., and Lindner, A., Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain*, 133: 262–271, 2010.
- [Voss *et al.*, 2010] Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., and Haggard, P., Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain*, 133: 3104–3112, 2010.
- [Wolpert, 1997] Wolpert, D.M., Computational approaches to motor control. *Trends in Cognitive Sciences*, 1: 209-216, 1997.
- [Treur and Umair, 2011] Treur, J., and Umair, M., A Cognitive Agent Model Using Inverse Mirroring for False Attribution of Own Actions to Other Agents. In: *Proc. of the 24th Intern. Conf. on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE'11*. Lecture Notes in Artificial Intelligence, Springer Verlag, 2011.