A Model-Based Ambient Agent Providing Support in Handling Desire and Temptation

Mark Hoogendoorn[1], Zulfiqar A. Memon[2], Jan Treur[1], Muhammad Umair[3]

[1] VU University, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands.
[2]Sukkur Institute of Business Administration (Sukkur IBA) Air Port Road Sukkur, Pakistan
[3]COMSATS Institute of IT, Department of Computer Science Lahore, Pakistan
{mhoogen, zamemon, treur, mumair}@few.vu.nl

Abstract. An ambient agent system is presented estimating a human's dynamics of desiring and being tempted. The agent is equipped with a dynamical model of the human's processes which describes how a desire relates to responses in the form of being prepared for certain actions, which in turn relate to feelings which can be biased, due to experiences in the past. It is shown how by use of this dynamical model, the ambient agent is able to predict and assess a human's desire state, and his or her preparation for certain actions, and use this assessment to suggest alternatives to avoid falling for certain temptations.

Keywords: Ambient Agent, Dynamical Model, Desire, Temptation.

1 Introduction

Agent modelling provides a useful design approach to the area of ambient systems [1,18]. One of the more ambitious challenges in this area is to create ambient systems with an appropriate human-awareness: awareness of the (mental) states of humans. Human-aware ambient systems can be taken to perform a certain type of mindreading or to possess what in the psychological and philosophical literature is called a Theory of Mind [12, 14]. During the evolutionary human history, mindreading has been developed to address different types of mental states, such as desire, intention, attention, belief or emotion states [12]. Inspired by such capabilities as developed in nature, ambient systems can be designed that have mindreading capabilities for one or some of these types of mental states. Such mental states can be dynamic and often interact with each other. To obtain an adequate human-aware ambient system, a dynamical model describing these dynamics and interaction is needed. To design an ambient system incorporating such a model, agent modelling offers a useful approach, as agents are able to integrate such dynamical models and reason about them [4]. Human-aware ambient agents equipped with the ability to reason about the different types of mental states can be applied to support of humans, for example persons vulnerable to temptations due to a developing addiction. More specifically, this paper focuses on the dynamics and interaction of an individual's desires and temptations and integrates a domain model for these dynamics in an ambient agent model to provide effective support by an enhanced awareness of the cognitive and affective states of the person. A desire triggers a number of responses in the form of preparations for certain actions related to the desire that result in certain feelings. In a reciprocal manner, the generated feelings affect the preparations; for some literature on such reciprocal interactions between cognitive and affective states, see, for example, [11, 17, 19].

In this paper, first in Section 2 the domain model for the dynamics of desires, preparations and feelings is described. Section 3 presents the ambient agent model which integrates the domain model. Section 4 presents some simulation results of the integrated ambient agent model. In Section 5 verification of the integrated model is addressed. Section 6 is a discussion.

2 Desires, Preparations and Feelings

Any mental state in a person induces emotions felt by this person, as described in [10]. Following [9, 10] it is assumed that responses in relation to a mental state of desiring roughly proceed according to the following causal chain: desire $\rightarrow$ preparation for response $\rightarrow$ body state modification $\rightarrow$ sensing body state $\rightarrow$ sensory representation of body state $\rightarrow$ induced feeling. An 'as-if body loop' uses a direct causal relation preparation for response $\rightarrow$ sensory representation of body state as a shortcut in the causal chain; cf. [9]. The body loop (or as-if body loop) is extended to a *recursive (as-if) body loop* by assuming that the preparation of the bodily response is also affected by the state of feeling the emotion: feeling $\rightarrow$ preparation for the bodily response. Such recursion is suggested in [10], noticing that what is felt is a body state which is under control of the person. Within the model used in this paper, both the bodily response and the feeling are assigned a level or gradation, expressed by a number. The activation of a specific action preparation is based on both the activation level of the desire and of the feeling associated to this action. This illustrates Damasio's theory on decision making with a central role for emotions felt, called the Somatic Marker Hypothesis [2,7, 8]. Based on the recursive *as-if body loop*, not only the strength of the connection from desire to preparation but also the strength of the connection from feeling to preparation will play an important role in deciding which action to pursue. When one or each of these connections is weak it will not lead to a high activation level of the preparation state, whereas a strong connection strength may result in a high activation level of the preparation state so that it may become the dominant option that can play the role of a strong temptation.

The strengths of the connections from feeling to preparation are subject to learning. Especially when a specific action is performed and it leads to a strong effect in feeling, by Hebbian learning [3, 15] this may give a positive effect on the strength of this connection and consequently on future activations of the preparation of this specific action. Through such a mechanism experiences in the past may have their effect on behavioural choices made in the future. The ambient agent uses the model and hence, it is expressed in the next section.

3 The Ambient Agent Model

Based upon the domain model as briefly expressed in the previous section, and ambient agent model has been developed. The ambient agent model was specified in LEADSTO [6], in which both logical and numerical relations can be specified. Moreover in this language, direct temporal dependencies between two state properties in successive states are modeled by *executable dynamic properties.* The LEADSTO format is defined as follows. Let α and β be state properties of the form 'conjunction of ground atoms or negations of ground atoms'. In the LEADSTO language the notation $\alpha \twoheadrightarrow_{e, f, g, h} \beta$, means:

> *If state property α holds for a certain time interval with duration g,*
> *then after some delay (between e and f) state property β will hold*
> *for a certain time interval of length h.*

Here, atomic state properties can have a qualitative, logical format, such as an expression desire(d), expressing that desire d occurs, or a quantitative, numerical format such as an expression has_value(x, v) which expresses that variable x has value v. The LEADSTO language features a simulation tool that is able to execute the dynamic LEADSTO properties.

Within the ambient agent model, the model for the dynamics of desires, preparations and feelings as expressed in Section 2 was embedded in order to enable the agent to reason about this process, and to assess the person's desires, preparations and feelings. In psychology, this capability is often referred to as mindreading or Theory of Mind [12]. The embedding uses the format that the causal relationships of the model described in Section 2 above are transformed into relationships for beliefs of the ambient agent on mental states of the person. In order to achieve this, the idea of recursive modelling is used; e.g., [16]. This means that the beliefs that agents have about each other are represented in a nested manner. To this end, each mental state is parameterized with the name of the agent that is considered, thus creating concepts like has_state(human, feeling(b, 0.5)) , has_state(AA, performed(suggest(X)). In addition, a number of meta-representations are introduced. For example, has_state(AA, belief(has_state(human, feeling(b, 0.7)))) states that the ambient agent (AA) believes that the human has a feeling level of *0.7* for b. The following are the resulting agent local properties (ALP) that specify the processes within the ambient agent. The step size is indicated by Δt.The first properties specify how the agent AA observes the human's body state and creates a belief about it.

ALP1 Observing the human's body state
If the human has certain body state, then the ambient agent AA will observe this.
has_state(human, body_state(B, V, t))
$\rightarrow$ has_state(AA, observed(has_state(human, body_state(B, V, t))))

ALP2 Generating a belief for the human's body state
If the ambient agent AA observes that the human has certain body state,
then it will generate a belief on it.
has_state(AA, observed(has_state(human, body_state(B, V, t))))
$\rightarrow$ has_state(AA, belief(has_state(human, body_state(B, V, t))))

The desire considered in the example scenario is assumed to be generated by sensing an unbalance in a body state *b*, according to the principle that organisms aim at maintaining homeostasis of their internal milieu. The first dynamic property addresses how body states are sensed. The following properties specify how the ambient agent observes and generates beliefs about the human's sensing and sensory representation process.

ALP3 Generating a belief for a human's sensing

If AA believes that the human has certain body state,
then it will generate a belief that after Δt the human will sense this body state

has_state(AA, belief(has_state(human, body_state(B, V, t))))
$\rightarrow$ has_state(AA, belief(has_state(human, sensor_state(B, V, t+Δt))))

For the example scenario this dynamic property is used for *B* to estimate the person's sensing of the body state *b* from which the desire originates (e.g., a state of being hungry), and the body states b_i involved in feeling satisfaction with specific ways in which the desire is fulfilled. How sensory representations are generated is addressed in dynamic property ALP4.

ALP4 Generating a belief for the human's sensory representation

If AA believes that the human senses body state,
then it will generate a belief that after Δt the human will have a sensory representation for this.

has_state(AA, belief(has_state(human, sensor_state(B, V, t))))
$\rightarrow$ has_state(AA, belief(has_state(human, srs(B, V, t+Δt))))

A person's desire originates from the sensory representation of the body state unbalance. The ambient agent generates a belief on the human's desires by:

ALP5 Generating a belief for the human's desires

If AA believes that the human has a sensory representation for body state *b*
then it will generate a belief that after Δt the human will generate a desire

has_state(AA, belief(has_state(human, srs(b, V, t))))
$\rightarrow$ has_state(AA, belief(has_state(human, desire(b, V, t+Δt))))

Next it is shown how the ambient agent estimates the preparations that are triggered. It is assumed that within the person activation of a desire, together with a feeling, induces preparations for a number of action options: those actions the person considers relevant options to satisfy the desire, for example based on earlier experiences. Property ALP6 describes such responses to an activated desire in the form of the preparation for specific actions. It combines the activation levels V and V_i of two states (desire and feeling) through connection strengths ω_{1i} and ω_{2i} respectively. This specifies part of the recursive as-if loop between feeling and body state. This dynamic property uses a combination model based on a function $g(\sigma, \tau, V, V_i, \omega_{1i}, \omega_{2i})$ which includes a sigmoid threshold function

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

with steepness σ and threshold τ. For this model $g(\sigma, \tau, V, V_i, \omega_{1i}, \omega_{2i})$ is defined as $g(\sigma, \tau, V, V_i, \omega_{1i}, \omega_{2i}) = th(\sigma, \tau, \omega_{1i}V + \omega_{2i}V_i)$ with V, V_i activation levels and ω_{1i}, ω_{2i} weights of the connections to the preparation state.

ALP6 Generating a belief for the human's preparations

If AA believes that the human has a desire for *b* with level *V*
and AA believes that the human has feeling B_i with level V_i
and AA believes that the preparation for body state B_i has level U_i
and ω_{1i} is the strength of the connection from desire for *b* to preparation for B_i
and ω_{2i} is the strength of the connection from feeling of B_i to preparation for B_i
and σ_i is the steepness value for the preparation for B_i

and τ_i is the threshold value for the preparation for B_i
and γ_1 is the person's flexibility for bodily responses
then ambient agent AA will generate the belief that the human's preparation state for body state b_i will occur with level $U_i + \gamma_1(g(\sigma_i, \tau_i, V, V_i, \omega_{1i}, \omega_{2i}) - U_i)\, \Delta t$

has_state(AA, belief(has_state(human, desire(b, V, t)))) &
has_state(AA, belief(has_state(human, feeling(B, V_i, t)))) &
has_state(AA, belief(has_state(human, prep_state(B_i, U_i, t)))) &
has_steepness(prep_state(B_i), σ_i) & has_threshold(prep_state(B_i), τ_i)
→ has_state(AA, belief(
has_state(human, prep_state(B_i, $U_i + \gamma_1\, (g(\sigma_i, \tau_i, V, V_i, \omega_{1i}, \omega_{2i}) - U_i)\, \Delta t$), t+Δt)))

Variants of this property have been used to incorporate interventions which affect the preparations of some B_i : they are assumed to become *0* (suggestion not to do) or *1* (suggestion to do); for example:

has_state(AA, belief(has_state(human, desire(b, V, t)))) &
has_state(AA, belief(has_state(human, feeling(B_i, V_i, t))))
has_state(AA, belief(has_state(human, prep_state(B_i, U_i)))) &
has_state(human, sensor_state(suggestion(do, B_i))))) &
has_steepness(prep_state(B_i), σ_i) & has_threshold(prep_state(B_i), τ_i)
→ has_state(AA, belief(has_state(human, prep_state(B_i, 1, t+Δt))))

The following five properties describe how the ambient agent reasons about the human's body loop.

ALP7 Generating a belief for the human's sensory representation of body states
If AA believes that the human's preparation state for body state B_i with level V_1 occurred
and AA believes that the human senses his body state B_i with level V_2
and AA believes that the human has sensory representation for B_i with level U
and σ is the steepness value for the sensory representation for B_i
and τ is the threshold value for the sensory representation for B_i
and γ is the person's flexibility for bodily responses
then ambient agent AA will generate the belief that the human's sensory representation for body state B_i will occur with level $U + \gamma_2\, (g(\sigma, \tau, V_1, V_2, 1, 1) - U)\, \Delta t$.

has_state(AA, belief(has_state(human, prep_state(B_i, V_1, t)))) &
has_state(AA, belief(has_state(human, sensor_state(B_i, V_2, t))))
has_state(AA, belief(has_state(human, srs(B_i, U, t)))) &
has_steepness(srs(B_i), σ) & has_threshold(srs(B_i), τ)
→ has_state(AA, belief(has_state(human, srs((B_i, $U + \gamma_2\, (g(\sigma, \tau, V_1, V_2, 1, 1) - U)\, \Delta t$), t+Δt)

ALP8 Generating a belief for the human's feelings
If AA believes that the human has a sensory representation for body state B_i with level V,
then it will believe that the human has feeling B_i with level V.
has_state(AA, belief(has_state(human, srs(B_i, V, t)))) → has_state(AA, belief(has_state(human, feeling(B_i, V, t+Δt))))

Temporal relationships ALP9, ALP10 and ALP11 below describe the ambient agent's reasoning about how preparations of body states b_i and affect body states b and b_i. The idea is that the actions performed by body states b_i are different means to satisfy the desire related to b, by having an impact on the body state that decreases the activation level V (indicating the extent of unbalance) of body state b. In addition, when performed, each of them involves an effect on a specific body state b_i which can be interpreted as a basis for a form of satisfaction felt for the specific way in which b was satisfied. So, on the one hand a specific action performance involving b_i has an effect on body state b, by decreasing the level of unbalance entailed by b, and on the other hand it has an effect on the body state b_i

by increasing the level of satisfaction entailed by b_i. This level of satisfaction may be proportional to the extent to which the unbalance is reduced, but may also be disproportional.

As the possible actions to fulfill a desire are considered different, they differ in the extents of their effects on these two types of body states, according to an effectiveness rate α_i between 0 and 1 for b, and an effectiveness rate β_i between 0 and 1 for b_i. The effectiveness rates α_i and β_i can be considered a kind of connection strengths from the effector state to the body states b and b_i, respectively. In common situations for each action these two rates may be equal (i.e., $\alpha_i = \beta_i$), but especially in more pathological cases they may also have different values where the satisfaction felt based on rate β_i for b_i may be disproportionally higher or lower in comparison to the effect on b based on rate α_i (i.e., $\beta_i > \alpha_i$ or $\beta_i < \alpha_i$). An example of this situation would be a case of addiction for one of the actions. To express the extent of disproportionality between β_i and α_i, a parameter λ_i, called *satisfaction disproportion rate*, between -1 and 1 is used. This parameter relates β_i to α_i using a function f, by $\beta_i = f(\lambda_i, \alpha_i)$. Here the function $f(\lambda, \alpha)$ satisfies $f(0, \alpha) = \alpha$, $f(-1, \alpha) = 0$, and $f(1, \alpha) = 1$. The function $f(\lambda, \alpha)$ can be defined in a continuous (but not differentiable) manner as a piecewise linear function in λ by $f(\lambda, \alpha) = \alpha + \lambda(1-\alpha)$ if $\lambda \geq 0$, and $f(\lambda, \alpha) = (1+\lambda)\alpha$ if $\lambda \leq 0$. Using such f, for normal cases $\lambda_i = 0$ is taken, for cases were satisfaction is disproportionally higher $0 < \lambda_i \leq 1$ and for cases where satisfaction is disproportionally lower $-1 \leq \lambda_i < 0$.

ALP9 Generating a belief for the human's body modification
If AA believes that the human's preparation state for body state B_i with level V occurred,
then it will believe that the human's body state B_i is modified with level V.
has_state(AA, belief(has_state(human, prep_state(B_i, V, t))))
$\rightarrow$ has_state(AA, belief(has_state(human, effector_state(B_i, V, t+Δt))))

ALP10 Generating a belief for the human from effector state to modified body state b_i
If AA believes that the human's body B_i is modified with level V_i,
and AA believes that for each i the effectivity of B_i for b is α_i
and AA believes that the satisfaction disproportion rate of B_i for b is λ_i
then AA will believe that the human's body state B_i will have level $f(\lambda_i, \alpha_i)V_i$.
has_state(AA, belief(has_state(human, effector_state(B_i, V_i, t)))) &
has_state(AA, belief(is_effectivity_for(α_i, B_i, b))) &
has_state(AA, belief(is_disproportion_rate_for(λ_i, B_i)))
$\rightarrow$ has_state(AA, belief(has_state(human, body_state(B_i, f(λ_i, α_i)V_i), t+Δt)))

ALP11 Generating a belief for the human from effector state to modified body state b
If AA believes that the human's body B_i is modified with level V_i,
and AA believes that human's body state b has level V,
and AA believes that for each $_i$ the effectivity of B_i for b is α_i
then AA believes that human's body state b will have
level $V + (\vartheta * (1-V) - \rho * (1 - ((1 - \alpha_1 * V_1) * (1 - \alpha_2 * V_2) * (1 - \alpha_3 * V_3))) * V) \Delta t$.
has_state(AA, belief(has_state(human, effector_state(B_i, V_i, t)))) &
has_state(AA, belief(has_state(human, body_state(b, V, t)))) &
has_state(AA, belief(is_effectivity_for(α_i, B_i , b)))
$\rightarrow$ has_state(AA, belief(has_state(human, body_state(b,
V + (ϑ * (1-V) – ρ * (1 – ((1 - α_1*V_1) * (1 - α_2*V_2) * (1 - α_3*V_3))) * V) Δt), t+Δt))

Note that in case only one action is performed (i.e., $V_j = 0$ for all $j \neq i$), the formula in ALP11 above reduces to $V + (\vartheta * (1\text{-}V) - \rho * \alpha_i * V_i * V)\, \Delta t$. In the formula ϑ is a rate of developing unbalance over time (for example, getting hungry), and ρ a rate of compensating for this unbalance. Note that the specific formula used here to adapt the level of *b* is meant as just an example. As no assumptions on body state *b* are made, this formula is meant as a stand-in for more realistic formulae that could be used for specific body states *b*. A variant of this property has been used to incorporate external events *p* that incidentally increases the level of the body state (such as exercising):

has_state(AA, belief(has_state(human, effector_state(B_i, V_i, t)))) &
has_state(AA, belief(has_state(human, body_state(b, V, t)))) &
has_state(AA, belief(is_effectivity_for(α_i, B_i , b))) & external_effect(p)
→ has_state(AA, belief(has_state(human, body_state(b,
V+ ((ϑ+p) * (1-V) – ρ * (1 – ((1 - α_1*V_1) * (1 - α_2*V_2) * (1 - α_3*V_3))) * V) Δt), t+Δt))

The strengths ω_{2i} of the connections from feeling b_i to preparation of b_i are considered to be subjected to learning. When an action involving b_i is performed and leads to a strong effect on b_i, by Hebbian learning [3, 15] this increases the strength of this connection. This is an adaptive mechanism that models how experiences in the past may have their effect on behavioural choices made in the future, as also described in Damasio's Somatic Marker Hypothesis [7, 8]. Within the model the strength ω_{2i} of the connection from feeling to preparation is adapted using the following Hebbian learning rule. It takes into account a maximal connection strength *1*, a learning rate η, and an extinction rate ζ. A similar Hebbian learning rule can be found in [13]. The agent AA generates beliefs about the connection strengths based on Hebbian learning:

ALP12 Generating a belief for the human's Hebbian learning

If AA believes that the connection from feeling B_i to preparation of B_i has strength ω_{2i}
and AA believes that human has feeling B_i with level V_{1i}
and AA believes that the human's preparation of B_i has level V_{2i}
and the learning rate from feeling B_i to preparation of B_i is η
and the extinction rate from feeling B_i to preparation of B_i is ζ
then after Δt AA will believe that the connection from feeling B_i to preparation of B_i will have strength $\omega_{2i} + (\eta V_{1i} V_{2i} (1 - \omega_{2i}) - \zeta \omega_{2i})\, \Delta t$.

has_state(AA, belief(has_connection_strength(feeling(B_i), preparation(B_i), ω_{2i}, t))) &
has_state(AA, belief(has_state(human, feeling(B_i, V_{1i}, t)))) &
has_state(AA, belief(has_state(human, prep_state(B_i, V_{2i}, t)))) &
has_learning_rate(feeling(B_i), preparation(B_i), η) &
has_extinction_rate(feeling(B_i), preparation(B_i), ζ)
→ has_state(AA, belief(has_connection_strength(
feeling(B_i), preparation(B_i), ω_{2i} + ($\eta V_{1i} V_{2i}$ (1 - ω_{2i}) - $\zeta\omega_{2i}$) Δt)), t+Δt)

Based on the beliefs about the human's states an assessment is made on the level of desire, as follows (where, for example $th = 0.7$):

ALP13 Assessment generation

If AA believes that the human has a desire at time *t* with level *V* higher than threshold *th1*,
then an assessment will be generated by AA that human will have a high desire of *b* at time *t*

has_state(AA, belief(has_state(human, desire(b, V, t)))) & V ≥ th1
→ has_state(AA, assessment(has_state(human, high_desire(b), t)))

The desire assessment is used to generate an intervention intention, whenever needed. This intention persists until the point in time at which the intervention has to be performed. **Table 1** shows the criteria used in the agent's decision process.

Table 1. Assessment criteria used by the ambient agent

	Preparation state level > 0.1	Preparation state level ≤ 0.1
Effectivity rate > 0.5	A good option considered by the human	A good option not considered by the human
Effectivity rate ≤ 0.5	A bad option considered by the human	A bad option not considered by the human

Here the human is assumed to consider an option if the level of the associated preparation state is predicted above a certain threshold, which in the example scenario is set to *0.1*, whereas the different options that are available are characterized as good or bad based on the values of the effectivity rates of those options higher or lower than 0.5. Properties ALP14a – ALP17 below accomplish this intervention strategy:

ALP14a Generation of intended intervention by ambient agent: positive suggestion

If AA has generated an assessment that human will have a high desire for *b* at time *t*
and AA has desire of human's wellbeing
and AA believes that the human's preparation of B_i has level V_i
and AA believes that for each *i* the effectivity of B_i for *b* is α_i
and $V_i < 0.1$ and $\alpha_i > 0.5$
then AA will intends to intervene the human at a later time *t* to suggest for doing B_i

has_state(AA, assessment(has_state(human, high_desire(b), t))) & has_state(AA, desire(wellbeing(human))) &
has_state(AA, belief(has_state(human, prep_state(B_i, V_i, t+20)))) & $V_i < 0.1$ &
has_state(AA, belief(is_effectivity_for(α_i, B_i, b))) & $\alpha_i > 0.5$
→ has_state(AA, intended_intervention_at(suggestion(human, do, B_i), t))

ALP14b Generation of intended intervention by ambient agent: negative suggestion

If AA has generated an assessment that human will have a high desire for *b* at time *t*
and AA has desire of human's wellbeing
and AA believes that the human's preparation of B_i has level V_i
and AA believes that for each *i* the effectivity of B_i for *b* is α_i
and $V_i > 0.1$ and $\alpha_i < 0.5$
then AA will intends to intervene the human at a later time *t* to suggest for not doing B_i

has_state(AA, assessment(has_state(human, high_desire(b), t))) & has_state(AA, desire(wellbeing(human))) &
has_state(AA, belief(has_state(human, prep_state(B_i, V_i, t+20)))) & $V_i > 0.1$ &
has_state(AA, belief(is_effectivity_for(α_i, B_i, b))) & $\alpha_i < 0.5$
→ has_state(AA, intended_intervention_at(suggestion(human, don't_do, B_i), t))

ALP15 Propagation of intended intervention by ambient agent

If AA intends to intervene the human at a later time *t1* to suggest for eating
and the current time is *t2* and t2 < t1
then the intended intervention by AA will persist

has_state(AA, intended_intervention_at(X, t1)) & current_time(t2) & t2 < t1 → has_state(AA, intended_intervention_at(X, t1))

Finally the intervention is performed:

ALP16 Intervention by Ambient Agent

If AA intends to intervene the human at a later time *t1* to suggest for eating

and the current time is *t2*
and t2 = t1-3
and AA does not observes the human in eating
then AA will suggest the human to eat

has_state(AA, intended_intervention_at(X, t1)) & current_time(t2) & t2 = t1 – 3 →
has_state(AA, performed(X))

ALP17 Human sensing of the action performed by agent
If AA suggests human to do X
then the human will sense this

has_state(AA, performed(suggestion(human, X, B)))
→ has_state(human, sensor_state(has_state(AA, performed(suggestion(human, X, B)))))

4 Simulation Results for the Ambient Agent Model

A number of simulations have been performed within the LEADSTO simulation environment [6]. The model was tested in a small scenario, involving an ambient agent and a human, indicated by AA and human, respectively. The example scenario taken here considers a person who is getting hungry which generates a desire to eat for which a number of options is available at that time. As the level of desire increases this makes the person more tempted to eat, and in particular to choose the option that is associated to the best feeling. As the domain model is integrated within the ambient agent, it can predict the human's desire level well in advance, and assesses the extent to which the human will consider the different options that are available to fulfill this desire.

Based on the criteria given in **Table 1** above, if the ambient agent predicts that the human will consider those options that are not effective for fulfilling the desire, then it will suggest not to choose them. Similarly, if the assessment process of the ambient agent determines any options that are quite effective for the human to choose, but the human will not consider those, then it will suggest the human to choose them. The scenario starts with some initial values of the human's desire and feeling levels, and then keeps on updating this, using the integrated model explained in Section 3. An example simulation trace (under fixed parameter settings) is illustrated in **Fig 1** and **2** (here the time delays within the temporal LEADSTO relations were taken 1 time unit).

In these figures, where time is on the horizontal axis, the upper part shows the time periods, in which the binary logical state properties hold (indicated by the dark lines); for example, has_state(AA, assessment(has_state(human, high_desire(b), 204))). Below this part, quantitative information is provided about the human's actual desire, preparation states, connection strength levels, levels of different body states and the ambient agent AA's prediction of these. Values for these levels for the different time periods are shown by the dark lines. Note that the scale on the vertical axis differs over the different graphs, and only a selection of the relevant state properties is shown.

For the example trace shown in **Fig 1** and **2**, for each *i* that represents an option, $\lambda_i = 0$ was taken, so in this example simulation the human is not developing an addiction to any option. Option 3 has the highest effectiveness rate, i.e. $\alpha_3 = 1$. Its value is substantially higher than the rates for the other two

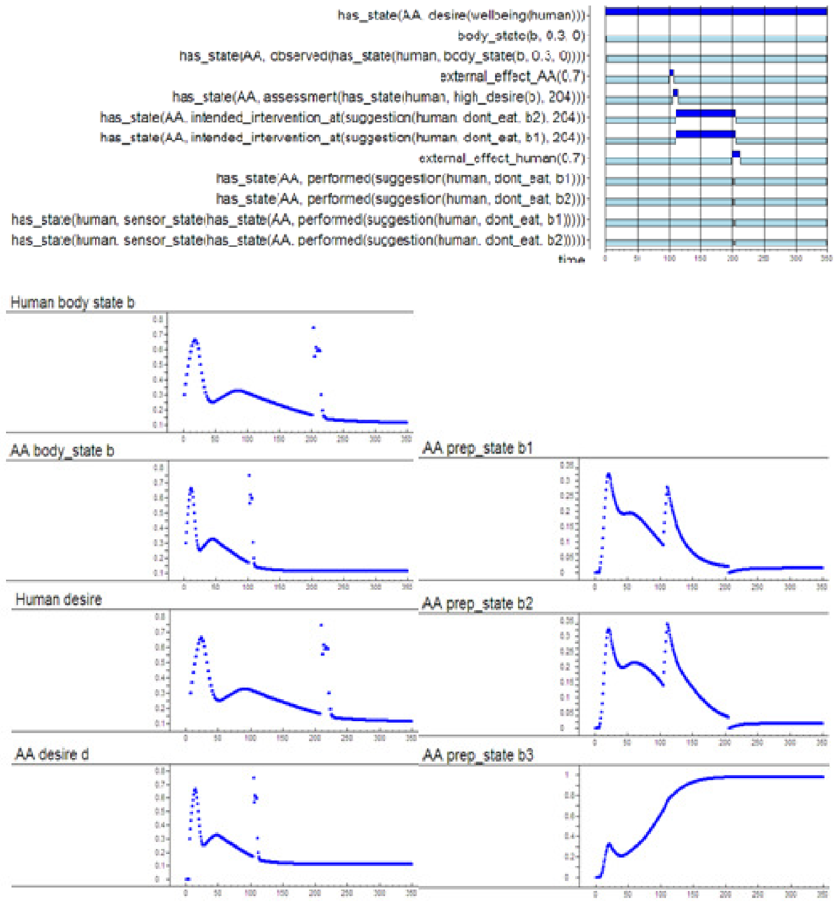

Fig 1 Simulation Trace 1 – Desire and preparation states ($\alpha_1=\beta_1=0.05$, $\alpha_2=\beta_2=0.25$, $\alpha_3=\beta_3=1$, $\gamma_1=\gamma_2=0.05$, $\sigma_1=\sigma_2=10$, $\tau_1=\tau_2=0.5$, $\rho=0.8$, $\vartheta=0.1$, $\eta=0.04$, $\zeta=0.01$)

available options. This affects the respective body states. Furthermore, as can be seen in **Fig 2** by the Hebbian learning it gives a strong effect on the strength of the connection from feeling to preparation for this option: the connection strength for option 3 increases over time until it reaches an equilibrium state.

As shown in the lower part of the **Fig 2**, at time point 10, the ambient agent predicts that the desire level of human will increase but it will not cross the threshold set to *0.7*, i.e., it is not considered sufficient enough to make the human tempted to choose this option. This is confirmed by the graph of the desire level of the human, where at time point 20, it increases but does not cross the threshold. Hence the ambient agent does not intend to perform any action. But later, some external effects (e.g., the human's habit to attend gym) causes an increase in this

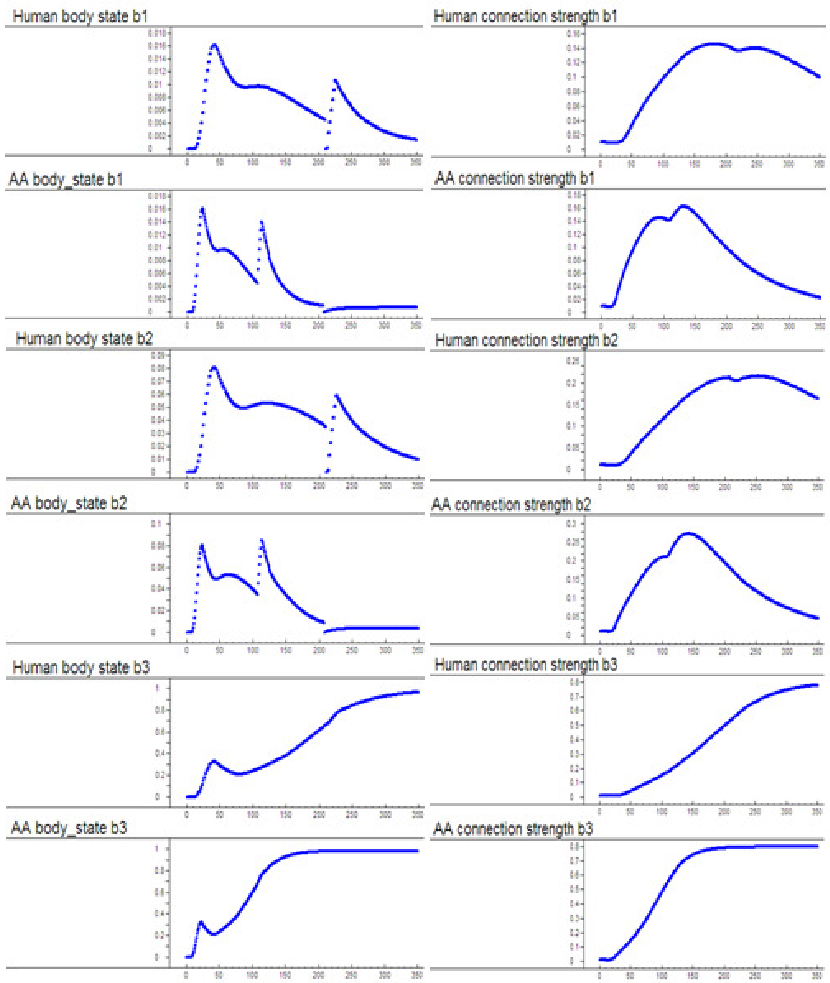

Fig 2 Simulation Trace 1 – Normal behavior: adaptation process

desire level, which is predicted by the ambient agent AA in the simulation at time point 102, as shown in the upper part of the **Fig 1**, by the state property has_state(AA, assessment(has_state(human, high_desire(b), 204))), expressing that an assessment has been generated that the human will have a high desire for b at time 204. Thereafter, as described in **Table 1**, AA predicts that the human will choose all three options because of the high values of the preparation states for those options, as shown in **Fig 1**, in the graph of AA prep_state b1, AA prep_state b2 and AA prep_state b3. After this, the agent will assess for these options whether they are good or bad, based on their effectivity rates. For this particular example simulation, the options *b1* and *b2* are assessed as bad because of their low effectivity rates, i.e., $\alpha_1 = 0.05$, $\alpha_2 = 0.25$, which are lower than the threshold set to 0.5. On the other hand, option *b3* is assessed as good because its effectivity rate is higher than threshold, i.e., $\alpha_3 = 1$. Hence the ambient agent generates the intention to suggest the human not to choose options *b1* and *b2* as shown in the upper part of **Fig 1**, by the state property has_state(AA, intended_intervention_at(suggestion(human, don't_eat, b1), 204)) and similarly for *b2*.

5 Analysis of the Ambient Agent Model by Automated Verification

In order to investigate whether the ambient agent indeed acts according to what is expected, some logical properties (requirements) have been identified, formalised, and verified against the simulation traces of the model. In this section, first the language used to express such properties is briefly introduced, followed by the specification of the actual properties, and the result of their verification. Using a formal specification for desired properties of the ambient agent enables automatic verification of them against simulation traces. This automated verification is performed using the logical language TTL and its software environment [5]. The temporal predicate logical language TTL supports formal specification and analysis of dynamic properties, covering both qualitative and quantitative aspects. TTL is built on atoms referring to *states* of the world, *time points* and *traces*, i.e. trajectories of states over time. *Dynamic properties* are temporal statements formulated with respect to traces based on the state ontology Ont in the following manner. Given a trace γ over state ontology Ont, the state in γ at time point t is denoted by state(γ, t). These states are related to state properties via the infix predicate |=, where state(γ, t) |= p denotes that state property p holds in trace γ at time t. Based on these statements, dynamic properties are formulated in a sorted first-order predicate logic, using quantifiers over time and traces and the usual first-order logical connectives such as $\neg, \wedge, \vee, \Rightarrow, \forall, \exists$. For more details, see [5]. An overall property to be satisfied by the agent is that if the level of a desire of the human exceeds a particular threshold, it should eventually become below the threshold.

P1(th:real): Successful support

In all traces γ and all time points t the level of a desire of the human if the desire of a human exceeds the threshold, then there exists a later time point at which this is not the case.

∀t:TIME, γ:TRACE, V:REAL

[state(γ, t) |= desire(b, V) & V > th ⇒ ∃t2:TIME, V2:REAL [state(γ, t2) |= desire(b, V2) & V2 < th]]

For the simulation traces generated using the ambient agent model, this property is satisfied for all traces (with a threshold value of *0.7*). The overall behavior as expressed in P1 can be accomplished by intervention by giving one or more suggestions at the right moment (expressed in P2) in combination with the human responding to these suggestions (expressed in P3).

P2(th:real, d:duration): Right moment for intervention

In all traces γ, if the ambient agent at time point t1 predicts that at time point t2 the human will have a desire exceeding the threshold th, then the ambient agent will give a suggestion to the human.

∀t1, t2:TIME, γ:TRACE, V:REAL
[[state(γ, t1) |= has_state(AA, belief(has_state(human, desire(b, V, t2)))) & V > th]
⇒ ∃t3:TIME > t1:TIME, A:ACTION, B:BODY_STATE
[state(γ, t3) |= has_state(AA, performed(suggestion(human, A, B)))]]

This property holds for all traces (when a threshold of *0.7* is chosen).

P3(d:duration): Right response

In all traces γ, if the ambient agent gives a suggestion to the human at time point t to either avoid a body state *B* (don't eat for this case) or accomplish a body state *B* (i.e., eat), then the human will follow this suggestion, indicated by a preparation state for *B* being *0* for the case of an avoidance suggestion, or a *1* in case of an accomplish suggestion for the body state *B*.

∀t1:TIME, γ:TRACE, V:REAL, B:BODY_STATE
[[state(γ, t1) |= has_state(AA, performed(suggestion(human, dont_do, B))) ⇒
∃t2:TIME > t1 [state(γ, t2) |= prep_state(B, 0)]] &
[state(γ, t1) |= has_state(AA, performed(suggestion(human, do, B))) ⇒
∃t2:TIME > t1 [state(γ, t2) |= prep_state(B, 1)]]

This last property is satisfied for all traces as well.

6 Discussion

To function in a knowledgeable manner, ambient agents [1, 18] need a model of the humans they are supporting. Such a model enables them to perform a form of mindreading [12, 14]. The ambient agent model presented here focuses on mindreading concerning the interaction between desires, preparations and feelings, based on neurological theories that address this interaction. The integrated dynamical model describes more specifically how a desire induces (as a response) a set of preparations for a number of possible actions, involving certain body states, which each affect sensory representations of the body states involved and thus provide associated feelings. On their turn these feelings affect the preparations, for example, by amplifying them. In this way an agent model is obtained for desiring which integrates both cognitive and affective aspects of mental functioning. For the interaction between feeling and preparation of responses, a converging recursive body loop is included in the dynamical model, based on elements taken from [9, 10]. Both the strength of the preparation and of the feeling emerge as a result of the dynamic pattern generated by this loop. The dynamical model is adaptive in the sense that within these loops the connection strengths from feelings to preparations are adapted over time by Hebbian learning [3, 13, 15]. By this adaptation mechanism, in principle the person achieves that the most effective action to fulfill a desire is chosen. However, the dynamical model can also be used to cover humans for whom satisfaction for an action is not in proportion with the fulfillment of the desire, as occurs, for example, in certain cases of earlier addictive experiences which provide temptations for the future. In this case, action choice may become biased by such temptations, and this is where an ambient agent can play a supporting role. The agent model equipped with the dynamical model for the dynamics of desires, preparations and feelings was specified in the hybrid dynamic modelling language LEADSTO, and simulations were performed in its software environment [6]. Simulation experiments show that the model behaves as expected, which also have been verified formally.

References

1. Aarts, E.; Collier, R.; Loenen, E. van; Ruyter, B. de (eds.) (2003). Ambient Intelligence. Proc. of the First Eur. Symposium, EUSAI 2003. Lecture Notes in Computer Science, vol. 2875. Springer Verlag, 2003.
2. 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.
3. 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.
4. Bosse, T., Hoogendoorn, M., Klein, M., and Treur, J., (2009). A Generic Agent Architecture for Human-Aware Ambient Computing. In: Mangina, E., et al. (eds.), Agent-Based Ubiquitous Computing. Series on Ambient and Pervasive Intelligence, vol. 1. World Scientific Publishers: Atlantis Press, 2009, pp. 35-62.
5. Bosse, T., Jonker, C.M., Meij, L. van der, Sharpanskykh, A., and Treur, J., (2009). Specification and Verification of Dynamics in Agent Models. In: Int. Journal of Cooperative Information Systems, vol. 18, 2009, pp. 167-193.
6. 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.
7. Damasio, A. (1994). Descartes' Error: Emotion, Reason and the Human Brain. Papermac, London.
8. Damasio, A. (1996). The Somatic Marker Hypothesis and the Possible Functions of the Prefrontal Cortex. *Philosophical Trans. of the Royal Society: Biological Sciences*, vol. 351, pp. 1413-1420
9. Damasio, A. (1999). The Feeling of What Happens. Body and Emotion in the Making of Consciousness. New York: Harcourt Brace, 1999.
10. Damasio, A. (2003). Looking for Spinoza. Vintage books, London, 2004.
11. Eich, E., Kihlstrom, J.F., Bower, G.H., Forgas, J.P., & Niedenthal, P.M. (2000). Cognition and Emotion. New York: Oxford University Press.
12. Gärdenfors, P. (2003). How Homo Became Sapiens: On The Evolution Of Thinking. Oxford University Press, 2003.
13. Gerstner, W., and Kirstner, W.M. (2002). Mathematical formulations of Hebbian learning. Biol. Cybern., vol. 87, pp. 404-145
14. Goldman, A.I. (2006). Simulating Minds: the Philosophy, Psychology and Neuroscience of Mindreading. Oxford University Press.
15. Hebb, D. (1949). The Organisation of Behavior. Wiley, New York.
16. Marsella, S.C., Pynadath, D.V., and Read, S.J. (2004). PsychSim: Agent-based modeling of social interaction and influence. In: Lovett, M. et al. (eds.), Proc. of the Int. Conf. on Cognitive Modelling, ICCM'04, pp. 243-248.
17. Niedenthal, P.M. (2007). Embodying Emotion. Science, vol. 316, (2007), pp. 1002-1005.
18. Riva, G., F. Vatalaro, F. Davide, M. Alcañiz (eds.) (2005). Ambient Intelligence. IOS Press.
19. Winkielman, P., Niedenthal, P.M., and Oberman, L.M. (2009). Embodied Perspective on Emotion-Cognition Interactions. In: Pineda, J.A. (ed.), Mirror Neuron Systems: the Role of Mirroring Processes in Social Cognition. Springer Science, 2009, pp. 235-257.