A Cognitive Architecture for Mental Processes Involving Mental Models Analysed from a Self-modeling Network Viewpoint

van Ments, Laila; Bhalwankar, Raj; Treur, Jan

published in
Data Science and Intelligent Systems
2021

DOI (link to publisher)
10.1007/978-3-030-90321-3_52

document version
Publisher's PDF, also known as Version of record

A Cognitive Architecture for Mental Processes Involving Mental Models Analysed from a Self-modeling Network Viewpoint

Laila van Ments[1], Raj Bhalwankar[2], and Jan Treur[2](✉)

[1] AutoLeadStar, Jerusalem, Israel
laila@autoleadstar.com

[2] Social AI Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
r.p.bhalwankar@student.vu.nl, j.treur@vu.nl

Abstract. This paper contributes an analysis of how in mental and social processes, humans often apply specific mental models and learn and adapt them in a controlled manner. It is discussed how controlled adaptation relates to the Plasticity Versus Stability Conundrum in neuroscience. From the analysis an informal three-level cognitive architecture for controlled adaptation was obtained. It is discussed here from a self-modeling network viewpoint how this cognitive architecture can be modeled as a self-modeling network. Making use of the specific network characteristics offered by the self-modeling network structure format, a large number of options for different types of adaptation of mental models and different types of control over adaptation of mental models were obtained. Many of these options were illustrated by a several realistic examples that were formalized by self-modeling networks. Other options that were distinguished from the analysis here, are offered as interesting options for future research.

1 Introduction

The area of mental models within psychology, educational science and other related disciplines addresses how in their mental and social processes, humans often learn, adapt and apply specific mental models as a kind of blueprints, schemas or maps. Most of the many publications on mental models in multiple disciplines are informal and not computational. This may partly be due to the challenging complexity of the different types of processes involved in handling mental models. The main processes that came out of a more detailed analysis of this literature are (Van Ments and Treur 2021c):

(1) *Applying a mental model*
This can be considered a form of internal (mental) simulation. Outcomes of this, affect a person's decisions and actions; e.g., (Craik 1943)

(2) *Developing and maintaining a mental model*
Adaptation of mental models often takes place. This usually involves learning, extinction or forgetting, and revision; e.g., (Piaget 1936; Hebb 1949; Seel 2006)

R. Silhavy et al. (Eds.): CoMeSySo 2021, LNNS 231, pp. 628–647, 2021.
https://doi.org/10.1007/978-3-030-90321-3_52

(3) *Exerting control over a mental model*

In a context-sensitive manner, usually control is exerted over adaptation of a mental model; e.g., (Du Plooy 2016; Darling-Hammond et al. 2008; Hurley 2008; Mahdavi 2014; Pintrich 2000).

To obtain a formalized and computational model of mental processes involving a mental model, these three different types of interacting processes all have to be addressed, which indeed may be a bit challenging.

Some inspiration can be obtained from the wider neuroscientific context. In that context, (2) corresponds to what often is called plasticity and (3) relates to the notion of metaplasticity. In (Sjöström et al. 2008), the latter topic is discussed in relation to what is called the Plasticity Versus Stability Conundrum. More specifically, concerning (2) and (3), within neuroscience it has been found more in general that:

- In the brain *plasticity* can occur in different forms; for example:
  - synaptic neural plasticity; e.g., Hebbian learning (Hebb 1949)
  - nonsynaptic neural plasticity (sometimes called intrinsic plasticity) such as plasticity of excitability thresholds within neurons; e.g., (Chandra and Barkai 2018; Debanne et al. 2019; Sjöström et al. 2008)
- Plasticity turns out not to be constant but can be depend on circumstances; various neural mechanisms have been discovered by which the extent of plasticity varies over different circumstances by being controlled in a context-sensitive manner. This is called *metaplasticity*; e.g., (Abraham and Bear 1996; Magerl et al. 2018; Robinson et al. 2016; Sjöström et al. 2008)

These concepts and the way in which they have been modeled by formalized and computational models in (Treur 2020) provided useful inspiration to obtain formalised computational models for mental processes based on mental models as well.

As a first step, based on the analysis of the different types of interacting processes (1), (2) and (3) involved in mental model handling, an informal global cognitive architecture for mental model handling has been developed taking these processes into account in three different but related levels; see (Van Ments and Treur 2021c). Moreover, using the viewpoint of self-modeling networks, it has been shown how this (informal) cognitive architecture can be formalized in a computational manner in a self-modeling network format. In this self-modeling network, for the abovementioned informal three-level cognitive architecture, by the base level network internal simulation based on a mental model as subnetwork takes place, by a first-order self-model network at the next level, adaptation of this subnetwork representing the mental model and by a second-order self-model network at the second-next level context-sensitive control over this adaptation takes place. This will be discussed in some more detail in Sect. 3. After that, in Sects. 4, 5 and 6 each of the three levels will be discussed in more detail and illustrated by many examples of realistic cases of mental processes involving mental models. First, in Sect. 2 the notion of self-modeling network model is briefly introduced.

2 Self-modeling Network Models

A specific modeling approach addressing dynamics and adaptivity is the network-oriented modeling approach described in (Treur 2020). The current section briefly describes this modeling approach.

2.1 Network Models

According to the network-oriented modeling approach described in (Treur 2020) a network model is characterised by:

- **connectivity characteristics**
  Connections from a node (or state) X to a node Y and their *weights* $\boldsymbol{\omega}_{X,Y}$
- **aggregation characteristics**
  For any node Y, some *combination function* $\mathbf{c}_Y$**(..)** defines aggregation that is applied to the single impacts $\boldsymbol{\omega}_{X,Y}X(t)$ on Y through its incoming connections from states X
- **timing characteristics**
  Each node Y has a *speed factor* $\boldsymbol{\eta}_Y$ defining how fast it changes for given (aggregated) impact

The difference (or differential) equations that are useful for simulation purposes and also for analysis of network dynamics incorporate these network characteristics $\boldsymbol{\omega}_{X,Y}$, $\mathbf{c}_Y$(..), $\boldsymbol{\eta}_Y$: it holds.

$$Y(t + \Delta t) = Y(t) + \boldsymbol{\eta}_Y\left[\mathbf{c}_Y\left(\boldsymbol{\omega}_{X_1,Y}X_1(t), \ldots, \boldsymbol{\omega}_{X_k,Y}X_k(t)\right) - Y(t)\right]\Delta t \tag{1}$$

for any state Y and where $X_1, \ldots, X_k$ are the states from which it gets its incoming connections. Examples of useful combination functions are:

- the simple logistic sum function $\mathbf{slogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}$(..) defined by:

$$\mathbf{slogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, \ldots, V_k) = \frac{1}{1 + \mathrm{e}^{-\boldsymbol{\sigma}(V_1+\ldots+V_k-\boldsymbol{\tau})}} \tag{2}$$

- the advanced logistic sum function $\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}$(..) defined by:

$$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, \ldots, V_k) = \left[\frac{1}{1 + \mathrm{e}^{-\boldsymbol{\sigma}(V_1+\ldots+V_k-\boldsymbol{\tau})}} - \frac{1}{1 + \mathrm{e}^{\sigma\tau})}\right]\left(1 + \mathrm{e}^{-\sigma\tau}\right) \tag{3}$$

  This function is obtained from the simple logistic sum function by subtracting its value for sum 0 from it and rescaling the result for the [0, 1] interval.

The above concepts enable to design network models and their dynamics in a declarative manner, based on mathematically defined functions and relations.

2.2 Modeling Adaptive Networks as Self-modeling Networks

Realistic network models are usually adaptive: their network characteristics often are adapted over time. Therefore, their dynamics is usually an interaction (sometimes called co-evolution) of these two sorts of dynamics: dynamics of the nodes (or states) in the network (dynamics *within* the network) versus dynamics of the characteristics of the network (dynamics *of* the network). Dynamics of the network's nodes are modeled declaratively by declarative mathematical functions and relations. In contrast, the dynamics of the network characteristics traditionally are described in a procedural, algorithmic nondeclarative manner, which then leads to a hybrid type of model. But by using *self-models* within the network, a network-oriented conceptualisation can also be applied to *adaptive* networks to obtain a declarative description using mathematically defined functions and relations; see (Treur 2020). This works through the addition of new nodes to the network (called *self-model states* or *reification states*) which represent (adaptive) network characteristics. Such nodes are depicted at a next level (*self-model level*), where the original network is at a *base level*. These types of characteristics with their self-model states and their roles are shown in Table 1.

This provides an extended network, also called *self-modeling network*. Like for all network models, a self-modeling network model is specified in a (network-oriented) declarative mathematical manner based on nodes and connections. These include interlevel connections relating nodes at one level to nodes on the other.

Table 1. Different network characteristics and self-model states for them

Types of characteristics	Concepts	Notations	Self-model states	Role played by the self-model state
Connectivity characteristics	Connections weights	$\boldsymbol{\omega}_{X,Y}$	$\mathbf{W}_{X,Y}$	Connection weight **W**
Aggregation characteristics	Combination functions and their parameters	$\mathbf{c}_Y(..)$ $\boldsymbol{\pi}_{i,j,Y}$	$\mathbf{C}_{i,Y}$ $\mathbf{P}_{i,j,Y}$	Combination function weight **C** Combination function parameter **P**
Timing characteristics	Speed factors	$\boldsymbol{\eta}_Y$	$\mathbf{H}_Y$	Speed factor **H**

The outcome is also a network model (Treur 2020, Chap. 10). This whole construction can be applied iteratively to obtain multiple self-model levels that can provide higher-order adaptive networks, and is quite useful to model, for example, plasticity and metaplasticity in the form of a second-order adaptive network with three levels, one base level and a first- and a second-order self-model level; e.g., (Treur 2020, Chap. 4).

To support the design of network models and simulation of them, for any application from a library predefined basic combination functions $\mathrm{bcf}_i(..)$, $i = 1, ..., m$ are selected by assigning weights $\boldsymbol{\gamma}_{i,Y}$, where the combination function then becomes the weighted average.

$$\mathbf{c}_Y(..) = (\boldsymbol{\gamma}_{1,Y}\mathrm{bcf}_1(..) + ... + \boldsymbol{\gamma}_{m,Y}\mathrm{bcf}_\mathrm{m}(..))/(\boldsymbol{\gamma}_{1,Y} + ... + \boldsymbol{\gamma}_{m,Y}) \tag{4}$$

Furthermore, parameters of combination functions are specified, so that $\mathrm{bcf}_i(..) = \mathrm{bcf}_i(\mathbf{p},\mathbf{v})$ where **p** is a list of parameters and **v** is a list of values.

3 Modeling the Cognitive Architecture for Mental Models as a Self-modeling Network

Based on the different processes in which mental models are used as briefly discussed above, a cognitive architecture for handling mental models has been designed covering the three types of processes in an integrated manner as depicted in Fig. 1, left hand side. For more details of this architecture, see (Van Ments and Treur 2021c).

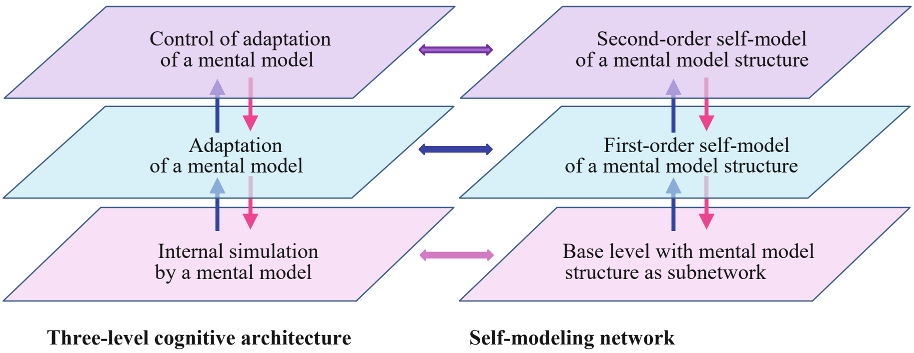

Fig. 1. Modeling the three-level cognitive architecture for mental model handling by a self-modeling network

The mapping from the three levels of the cognitive architecture to a self-modeling network is as follows (see also Fig. 1):

- **Lower level: a mental model as a subnetwork**
  The mental model as a relational structure at the base level within the cognitive architecture is modeled as a (sub)network structure of states (nodes) and connections between them at the base level of the self-modeling network; the dynamics of the states of this subnetwork model internal simulation of the mental model
- **Middle level: a first-order self-model of a mental model representing adaptation of its network structure**
  The level for adaptation of a mental model within the cognitive architecture is modeled as a first-order self-model of the mental model structure as represented at the base level; the dynamics of the states of this first-order self-model model adaptation by making changes in the structure of the mental model

- **Upper level: a second-order self-model of a mental model representing control of adaptation of its network structure**
  The level for control of adaptation of a mental model is modeled as a second-order self-model of the mental model, which is a self-model for the self-model for adaptation of the mental model; the dynamics of the states of this second-order self-model model control of adaptation by making changes in the structure of the first-order self-model that describes the adaptation of the mental model

So, mental models and the way they are handled can be considered as being described through multiple representations: they can be viewed from three levels of representation according to the three planes depicted in Fig. 1, right hand side. At the lower, base level depicted by the lower (pink) plane, a mental model, which in general essentially is considered to be a relational structure, is represented by nodes and connections between these nodes. For internal simulation, the nodes have activation levels that vary over time; based on the relations these activation levels affect each other over time. Next, at the adaptation level depicted in Fig. 1 right-hand side by the middle (blue) plane, it is represented how the mental model relations change over time by some adaptation specification. Finally, at the top level depicted by the upper (purple) plane in Fig. 1 it is indicated how the adaptation at the middle level is controlled. In this way, to model mental processes in which mental models play a role, within the self-modeling network these mental models do not get a single but a three-fold representation by which the different uses and operations on the mental model are distinguished like they are distinguished by the levels in the cognitive architecture.

In the next three sections each of the levels is discussed in some more detail and illustrated by many examples of realistic cases involving mental models in mental processes.

4 How Mental Models Can Be Used

From the viewpoint of the self-modeling network format, mental models are represented as subnetworks at the base level, which have their own internal connections (between their own mental model states), but also connections from and to other mental states that do not belong to the mental model. It is through the latter types of connections that a mental model can affect a person's mental processes in a wider sense and, more specifically, their decisions and actions. It was pointed out, among others already by Craik (1943) that having mental models enables organisms to make better decisions on how to act, as future developments can be predicted by these mental models:

> 'If the organism carries a "small-scale model" of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilise the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it.' (Craik 1943, p. 61)

From an analysis of many cases, it was found that mental models can affect a wide variety of types of other mental states and behavioural actions. Mental states relating to goals and actions, preparation and ownership for actions, and action execution are among the mental states and behaviour that can be affected by mental models. However, also other types of mental states can be affected by mental models, such as emotional responses, awareness states, belief states, and states in other mental models. In Table 1, an overview can be found of several realistic examples of how mental models affect other (base level) mental states which are not part of the mental model; this indicates the way in which the mental models are used in the overall mental processes and behaviour. All of them have been formalized in computational self-modeling network format based on the cognitive architecture shown in Fig. 1; see the references in the table to papers where they are described in much more detail. In particular, the following cases are addressed in Table 2.

Case 1 Multiple mental models
At the first row a reference is made to a visualization case study where states of a geometric mental model affect states of an arithmetical mental model to reveal arithmetical relations.

Case 2 Flashbacks in PTSD
The second row refers to how for PTSD a mental model made of a traumatic event is sometimes triggered (flashback) and then in turn triggers stressful emotions and awareness of the traumatic event.

Case 3 Counterfactual thinking
In the third row a counterfactual thinking case is addressed, where based on mental models of different alternative scenarios, beliefs are revised.

Case 4 Self-interpretation in therapy
The fourth row refers to a realistic therapy case in which self-interpretation based on a mental model of a person's own functioning leads to stronger awareness and revised beliefs.

Case 5 Metaphors for joint decision making
In the fifth row mental models for competitive and cooperative metaphors for joint decision making are considered that affect action ownership states that form the basis of a decision.

Case 6 Mental God-model and empathy
The sixth row refers to a case study of mental God-models, where these mental models affect a number of other mental states concerning the person's own actions, goals and also emotions; for example empathic or disempathic actions are influenced by a mental model of an empathic or disempathic God.

Case 7 Mental attachment model
The seventh row addresses a case for Attachment Theory, where mental models of self and other developed based on a primary caregiver during childhood, affect preparation for actions with respect to significant others later in life.

Case 8 Shared mental model
Finally, in the last row a case study for shared mental models for hospital teamwork is addressed, where, like in row 5, these mental models affect action decisions via their ownership states.

All in all, the influences of mental models on a person's mental processes at the base level can be diverse.

Table 2. Overview of the example mental models and which other mental states and behaviour they affect

		Action preparation	Action ownership	Action execution	Goal	Emotion	Awareness	Belief	Other mental model
Multiple mental models	(Treur 2021a)								+
Flashbacks in PTSD	(Van Ments and Treur 2021b)					+	+		
Counterfactual thinking	(Bhalwankar and Treur 2021c)							+	+
Self-interpretation in therapy	(Treur and Glas 2021)						+	+	
Metaphors for joint decision making	(Van Ments and Treur 2021c)		+						
Mental God-model and empathy	(Van Ments et al. 2022)	+	+	+	+	+			
Mental attachment model	(Hermans et al. 2021)	+							
Shared mental model	(Van Ments et al. 2021)		+						

To illustrate this in some more detail, for Case 2 and 3 the conceptual connectivity view on the network model design is provided in Fig. 2 and 3. Notice the three levels as introduced in Fig. 1: the base level in pink, the first-order self-model level to allow for adaptation in blue and the second-order self-model level to control adaptation in purple. Again, more depth on the models can be found in the references indicated in Table 2.

In Fig. 2 sensing of a traumatic event te consisting of a sequence of phases or steps is modeled by sensor states ss_{te1}, ss_{te2}, ss_{te3}. For example, te1 or traumatic event step 1, is a potentially dangerous situation for a child you observe, the second step te2 is an action from your side with the intention to save the child from that situation and te3 is an unfortunate failure of your action such that the child actually gets hurt. During this traumatic event sequence, sensory representations srs_{te1}, srs_{te2}, srs_{te3} are activated, and by sensory preconditioning (Brogden 1947; Hall 1996) the connections between these sensory representations are learned through a Hebbian learning mechanism (Hebb 1949). By this learning process, the mental model of the traumatic event sequence is formed and represented by base states srs_{te1}, srs_{te2}, srs_{te3} and their connections (see the small pink parallelogram within the base plane in Fig. 1) with first-order self-model states $\mathbf{W}_{srs_{te1},srs_{te2}}$ and $\mathbf{W}_{srs_{te2},srs_{te3}}$. What can be seen in Fig. 2 is that the mental model states have outgoing base level connections to two other mental states that are not part of the mental model: to the awareness state as_{te} of the trauma and to the emotional response preparation state ps_b. This is indeed what is shown in the second row of Table 1.

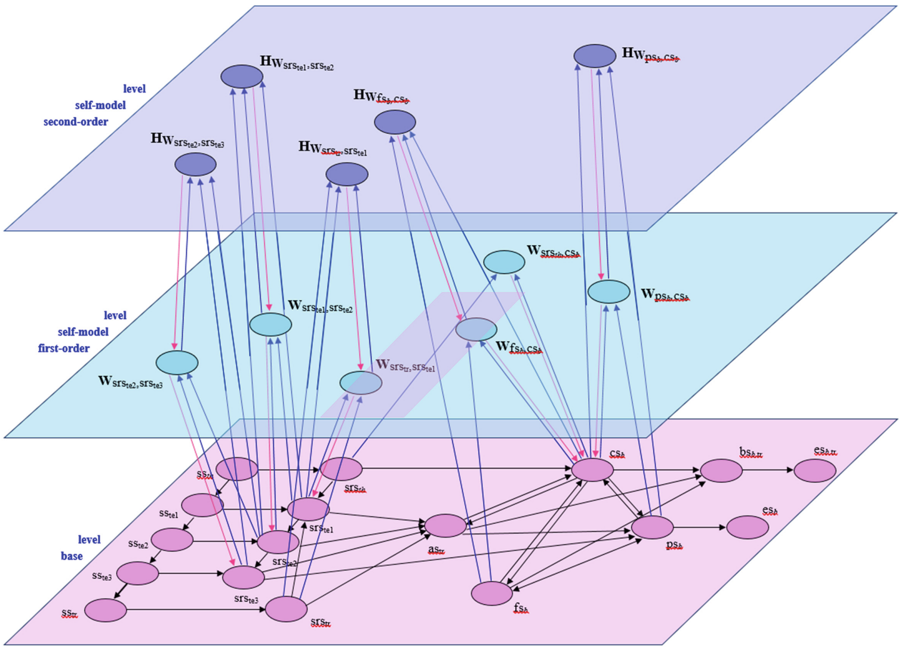

Fig. 2. Connectivity of the second-order adaptive network model of Case 2 described above

In the example for Case 3 shown in Fig. 3 the mental models are depicted within the small outlined areas within the (pink) base level plane. In this case, there is no direct connection within the base plane from mental model states to belief states but a causal pathway through the first-order self-model level (the middle blue plane); in this case (by a liberal interpretation) this is also considered as a causal effect on the belief, which at the same time is part of another mental model. Therefore, in row 3 of Table 1 there are two + indications.

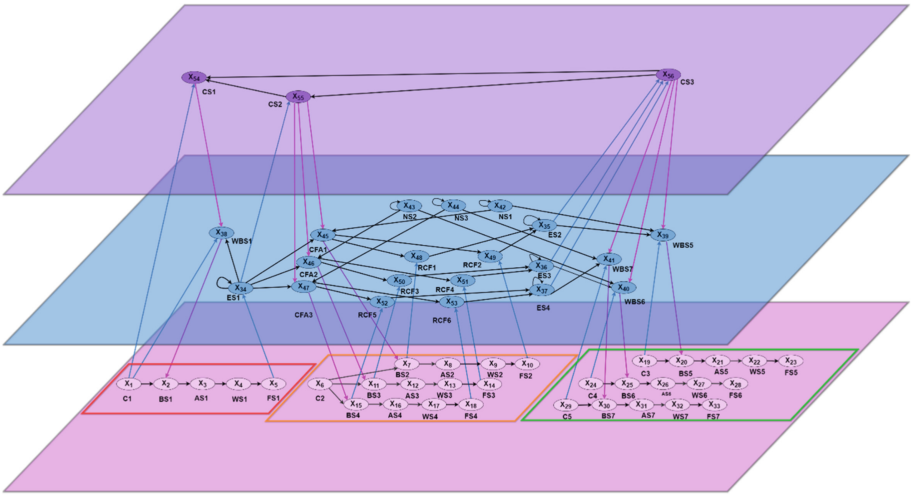

Fig. 3. Conceptual model of case 3 described above including the controlled adaptive network model for counterfactual thinking

5 How Mental Models Can Be Adapted

Different forms of adaptation or learning can be applied to mental models. Examples of individual types of learning are

- Learning of connections of mental models by observation of a process in the real world; for example, based on Hebbian learning (Hebb 1949)
- Learning of excitability of mental model nodes; for example, (Chandra and Barkai, 2018; Debanne, Inglebert, Russier, 2019; Sjöström et al. 2008)
- Learning of connections of mental models by instruction or by being told; for example by instructional learning, e.g., (Hogan and Pressley 1997; Seel 2006)
- Learning of connections of mental models from other, related mental models; for example, learning arithmetical or algebraic relations based on visualization like in Case 1 in Sect. 3; e.g., (Du Plooy 2016; Koedinger and Terao 2002)
- Learning of mental models based on counterfactual thinking like in Case 3 in Sect. 3; e.g., (Van Hoeck et al. 2015)
- Learning of mental models based on Theory of Mind self-interpretation, like in Case 4 in Sect. 3; for example (Treur and Glas 2021)

Also at the non-individual, social level, adaptation of mental models can play an important role. Three examples are

- Bonding based on homophily (or similarity), where the mental models describing the internal representations of connections between persons are adapted over time; e.g., (McPherson et al. 2001; Treur 2021b)
- Development of attachment relations as described by Attachment Theory, like in Case 7 in Sect. 3; e.g., (Bartholomew and Horowitz 1991; Hermans et al. 2021)

- Forgetting connections in a shared mental model during teamwork, like in Case 8 in Sect. 3; e.g., (Burtscher and Manser 2012; Van Ments et al. 2021)

From the viewpoint of the self-modeling network format as formalization, adaptation is described by (first-order) self-models of parts of the (base) network. This automatically offers a number of specific available network characteristics that are suitable for adaptation and that can be used in particular for adaptation of mental models represented as subnetworks. Examples of such networks characteristics that can be made adaptive are as follows.

Connectivity for mental model relations

- For example, mental model connections and their weights
- These can be modeled by self-model states $\mathbf{W}_{X,Y}$ representing the weight $\boldsymbol{\omega}_{X,Y}$ of the connection from mental model state X to mental model state Y

Aggregation for mental model states

- For example, excitability thresholds of mental model states
- These can be modeled by self-model states $\mathbf{T}_Y$ representing the excitability threshold $\boldsymbol{\tau}_Y$ of mental model state Y

Timing for mental model states

- For example, speed factors of mental model states
- These can be modeled by self-model states $\mathbf{H}_Y$ representing the speed factor $\boldsymbol{\eta}_Y$ of mental model state Y.

In Table 3, an overview can be found of such examples of adaptation of mental models that enabled to model a variety of realistic scenarios involving mental models. The first 8 rows correspond to the eight cases discussed in Sect. 3. In this table four new cases are shown:

Case 9 Car driving mental model

This case addresses the learning of a mental model of how a car works and how you can drive it; this takes place through a combination of instructional (by being told) and observational learning (by Hebbian learning).

Case 10 Self-controlled learning

In this case of a car driving mental model, the learner first learns by observation and then takes initiative to ask for (confirmative) instruction.

Case 11 Analysis and support
This case describes in the context of providing support, an adaptive mental model for analysis of possible problems and an adaptive mental model of suitable support actions for such problems.

Case 12 Bonding by homophily
This case addresses how the mental models of the (adaptive) connections between two persons are affected by how similar the (adaptive) mental models of characteristics or states of the two persons are.

For Case 2 depicted in Fig. 2 in row 2 of Table 2 it is indicated that the flashback movie is learned by Hebbian learning; for Case 3 depicted in Fig. 3, it is indicated in row 3 of Table 3 that it addresses learning by counterfactual thinking.

Table 3. Overview of realistic cases of how mental models can be adapted

		Learning by						Bonding by	
		Hebbian learning	Excitability learning	Information source	Counterfactual thinking	Other mental models	Theory of Mind analysis	Homophily	Attachment
Multiple mental models	(Treur 2021a)	+				+			
Flashbacks in PTSD	(Van Ments and Treur 2021a)	+							
Counterfactual thinking	(Bhalwankar and Treur 2021c)				+				
Self-interpretation in therapy	(Treur and Glas 2021)	+	+				+		
Metaphors for joint decision making	(Van Ments and Treur 2021b)	+							
Mental God-model and empathy	(Van Ments et al. 2022)	+							
Mental attachment model	(Hermans et al. 2021)	+							+
Shared mental model	(Van Ments et al. 2021)	+		+					
Car driving mental model	(Bhalwankar and Treur 2021a)	+		+					
Self-controlled learning	(Bhalwankar and Treur 2021b)	+		+					
Analysis and support	(Treur 2021c)	+	+						
Bonding by homophily	(Treur 2021b)							+	

6 How Mental Model Adaptation Can Be Controlled

In this section, different ways are discussed in which control over mental models and their adaptation can be exerted. Again this is done from the viewpoint of the network format provided by self-modeling networks. In (Sjöström et al. 2008), the topic of control of adaptation is discussed in relation to what is called the Plasticity Versus Stability Conundrum. Also within an AI-context, in machine learning examples of this conundrum and controlled adaptation to address it are known, such as the (decreasing) temperature parameter in simulated annealing and the sensitive balancing between exploration and exploitation in reinforcement learning, also called the explore-exploit dilemma (Holland 1975; March 1991; Wilson et al. 2014). This is explained in (Wilson et al. 2014) as follows:

> 'When you go to your favorite restaurant, do you always order the same thing, or do you try something new? Sticking with an old favorite ensures a good meal, but if you are willing to explore you might discover something better. This simple conundrum, deciding between something you know and something you do not, is commonly referred to as the exploration– exploitation dilemma.'

Applied to mental models in particular, this quote illustrates that on the one hand decision making based on known mental models can be very efficient (navigating based on a well-known map), but on the other hand this may prevent someone from learning even better decisions (exploring still unknown territory).

In a controlled adaptive network model for mental models based on the self-modeling network format, adaptation is modelled by a first-order self-model, as discussed in Sect. 4. There are a number of network characteristics involved in the structure of a first-order self-model used for the adaptation. By systematically going through these possible network characteristics, the following examples of network characteristics for adaptation to be controlled can be distinguished and are illustrated by various examples. Recall from Sect. 4 how exactly at the middle level self-model states can be introduced to represent adaptive network characteristics from the lower level. For example, self-model state $\mathbf{W}_{X,Y}$ represents an adaptive connection weight $\boldsymbol{\omega}_{X,Y}$ from the base level, and $\mathbf{H}_Y$ represents speed factor $\boldsymbol{\eta}_Y$ from the lower level, and so on. This can be iterated for the middle and upper level to obtain a second-order self-model. For example:

- second-order self-model state $\mathbf{H}_{\mathbf{W}_{\mathrm{X,Y}}}$ can be used to represent the adaptive adaptation speed of first-order self-model state $\mathbf{W}_{X,Y}$
- second-order self-model state $\mathbf{M}_{\mathbf{W}_{\mathrm{X,Y}}}$ can be used to represent the adaptive persistence parameter $\mu_{\mathbf{W}_{\mathrm{X,Y}}}$ of first-order self-model state $\mathbf{W}_{X,Y}$.
- second-order self-model state $\mathbf{W}_{Z,\mathbf{W}_{\mathrm{X,Y}}}$ can be used to represent the adaptive weight $\boldsymbol{\omega}_{Z,\mathbf{W}_{\mathrm{X,Y}}}$ of the connection from some state Z to state $\mathbf{W}_{X,Y}$.

For shortness, such second-order self-model states are sometimes called $\mathbf{H}_{\mathbf{W}}$-states, $\mathbf{M}_{\mathbf{W}}$-states, or $\mathbf{W}_{\mathbf{W}}$-states, whereas first-order self-model states can be called, for example, $\mathbf{W}$-states or $\mathbf{T}$-states. Following the different types of network characteristics used in the self-modeling network format, the following types of control can be distinguished.

- **Control by adaptive connectivity characteristics of first-order self-model states**
  - Adaptive connections of the causal pathways to the self-model states and their weights $\boldsymbol{\omega}$; for example:

    Choosing a mental model to be applied. For example, a decision to use a specific metaphor-based mental model as in the model in (Van Ments and Treur 2021b) or a decision to use a geometric mental model to support learning of an arithmetic mental model, as described in (Treur 2021a)
    Opening a communication channel from an information source to enable instructional learning of a mental model (decision to ask), as in the model in (Bhalwankar and Treur 2021b) and in the model described in (Treur 2021b)
    Opening an observation channel to enable observational learning of a mental model (decision to observe), as in the model described in (Treur 2021b)

  - Adaptive connections of the causal pathways from the self-model states to other states and their weights $\boldsymbol{\omega}$; for example:

    Modelling the effects of a chosen metaphor as in the model in (Van Ments and Treur 2021b)

- **Control by adaptive aggregation characteristics of first-order self-model states**
  - Adaptive choice of combination function; for example:

    For Hebbian learning of mental model connections a weighted average of $\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W)$ and $\mathbf{smin}_{\boldsymbol{\lambda}}(V_1, V_2)$, with adaptive weights $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$.

  - Adaptive parameters of chosen combination functions; for example:

    Adaptive values for the persistence factor $\boldsymbol{\mu}$ of $\mathbf{hebb}_{\boldsymbol{\mu}}(..)$ as in the self-modeling network model for shared mental models described in (Van Ments et al. 2021) or for the scaling factor $\boldsymbol{\lambda}$ of $\mathbf{smin}_{\boldsymbol{\lambda}}(V_1, V_2)$.

- **Control by adaptive timing characteristics of first-order self-model states**
  - Adaptive adaptation speed (learning rate) $\boldsymbol{\eta}$; for example:

    Addressing the Plasticity Versus Stability conundrum (Sjöström et al. 2008) based on some context factors indicating when plasticity is needed fully and when plasticity should be limited or frozen.
    Accelerating adaptation speed upon increased stimulus exposure (Robinson et al. 2016), for example as applied in the example model in (Van Ments and Treur 2021b)

As discussed above, in a self-modeling network format, control of any of such network characteristics (for first-order self-models for adaptation of a mental model) is modeled by a second-order self-model. To illustrate this, based on the above distinctions, in Table 3 a summarized overview is given of several cases of applications of second-order self-models to control adaptation of mental models as also collected in (Treur and Van Ments 2022).

For case 2 depicted in Fig. 2 in row 2 of Table 4 it is indicated that for the learning of the flashback movie by Hebbian learning the adaptation speed is controlled; for case 3 depicted in Fig. 3, it is indicated in row 3 of Table 4 that for the learning by counterfactual thinking the exchange between the different mental models is controlled.

Table 4. Overview of example mental models: how they are controlled

		Control of adaptation speed	Control of persistence	Control of communication	Control of observation	Control of exchange
Multiple mental models	(Treur 2021a)					+
Flashbacks in PTSD	(Van Ments and Treur 2021a)	+				
Counterfactual thinking	(Bhalwankar and Treur 2021c)					+
Self-interpretation in therapy	(Treur and Glas 2021)					+
Metaphors for joint decision making	(Van Ments and Treur 2021b)	+				
Mental God-model and empathy	(Van Ments et al. 2022)	+				
Mental attachment model	(Hermans et al. 2021)	+				
Shared mental model	(Van Ments et al. 2021)		+			
Self-controlled learning	(Bhalwankar and Treur 2021b)			+		
Analysis and support	(Treur 2021c)	+				+
Bonding by homophily	(Treur 2021b)			+	+	

Moreover, in Table 5, a more complete overview is obtained of different types of control against different types of learning. Note that in most cells in Table 5 further references are included, but for those cells where no references are included, in general this means that these options are yet to be explored in detail. This contributes to a future research agenda.

Table 5. Overview of different types of controlled adaptation for example mental models as collected in (Treur and Van Ments 2022)

Learning by	Control of adaptation via connectivity	Control of adaptation via aggregation	Control of adaptation via timing
Observation and monitoring	Controlled learning by observation for a mental model **W**-state for bonding via a $\mathbf{W_W}$-state (Treur and Van Ments, 2022, Chap. 13); formation of a mental model of another person; see also (Hermans et al. 2021)	Hebbian mental model learning **W**-state persistence control via an $\mathbf{M_W}$-state (Treur and Van Ments 2022, Chap. 14); controlled forgetting of a mental model relation; see also (Van Ments et al. 2021)	Hebbian mental model learning **W**-state speed control via an $\mathbf{H_W}$-state: adaptation accelerates with increasing exposure (Treur and Van Ments 2022, Chap. 5, Chap. 7, Chap. 10, Chap. 11, Chap. 12); e.g., learning mental models for flashback experiences (Van Ments and Treur 2021b), analysis and support tasks (Treur 2021c), metaphors (Van Ments and Treur 2021c), mental God-model (Van Ments et al. 2022), self- and other-models (Hermans et al. 2021)
Excitability adaptation	Incoming connection for an adaptive mental model excitability **T**-state control via a $\mathbf{W_T}$-state (Treur and Van Ments 2022, Chap. 7); learning excitability (Debanne et al. 2019; Chandra and Barkai 2018) of a mental model's states; see also (Treur 2021c)	Excitability mental model learning **T**-state aggregation control, for example through adaptive (steepness $\boldsymbol{\sigma}$ and threshold $\boldsymbol{\tau}$) parameters of a logistic combination function used for the **T**-state represented by $\mathbf{S_T}$ - and $\mathbf{T_T}$ - states	Excitability mental model learning **T**-state speed control via an $\mathbf{H_T}$-state (Treur and Van Ments 2022, Chap. 7); learning excitability (Debanne et al. 2019; Chandra and Barkai 2018) of a mental model's states; see also (Treur 2021c)

(*continued*)

Table 5. (*continued*)

Learning by	Control of adaptation via connectivity	Control of adaptation via aggregation	Control of adaptation via timing
Communication	Learner-controlled instructional learning of a mental model **W**-state via a $\mathbf{W_W}$-state (Treur and Van Ments 2022, Chap. 9); opening a communication channel with the instructor by asking; see also (Bhalwankar and Treur 2021b) Controlled learning by communication for a mental model **W**-state for bonding via a $\mathbf{W_W}$-state (Treur and Van Ments 2022, Chap. 13); opening a communication channel with the other person by asking; see also (Treur 2021b)	Learner-controlled instructional learning of a mental model **W**-state via a $\mathbf{T_W}$-state for excitability (Debanne et al. 2019; Chandra and Barkai 2018) of the **W**-state (opening a communication channel with the instructor by more sensitive listening)	Learner-controlled instructional learning of a mental model **W**-state via an $\mathbf{H_W}$-state (controlling the timing of a communication channel with the instructor)
Other mental models	Controlled connection **W**-state for counterfactual activation of a mental model via a $\mathbf{W_W}$-state (Treur and Van Ments, 2022, Chap. 6); see also (Bhalwankar and Treur 2021c)	Controlled connection **W**-state for counterfactual activation of a mental model via a $\mathbf{T_W}$-state addressing excitability (Debanne et al. 2019; Chandra and Barkai, 2018) of the **W**-state	Controlled inter mental model exchange connection **W**-state via $\mathbf{H_W}$-state (Treur and Van Ments, 2022, Chap. 4); exchange from arithmetic mental model to geometric mental model; see also (Treur 2021a)

(*continued*)

Table 5. (*continued*)

Learning by	Control of adaptation via connectivity	Control of adaptation via aggregation	Control of adaptation via timing
Contextual factors	Controlled connection **W**-state for activation of mental model via a $\mathbf{W_W}$-state (Treur and Van Ments 2022, Chap. 6); activation of a mental model for possible future action; see also (Bhalwankar and Treur 2021c)	Controlled connection **W**-state for activation of mental model via a $\mathbf{T_W}$-state addressing excitability of the **W**-state based on contextual factors	Controlled adaptive mental model effect connection **W**-state via an $\mathbf{H_W}$-state (Treur and Van Ments 2022, Chap. 10); adapting the own choices based on the context given by the other person; see also (Van Ments and Treur 2021c)

7 Discussion

This paper addressed an analysis of how in mental and social processes, humans often apply specific mental models and learn and adapt them in a controlled manner. Part of it is based on (Bhalwankar et al. 2021). It was discussed how controlled adaptation relates to the Plasticity Versus Stability Conundrum in neuroscience (Sjöström et al. 2008). From the analysis, an informal three-level cognitive architecture for controlled adaptation was obtained. It was discussed from a self-modeling network viewpoint how this cognitive architecture can be modeled as a self-modeling network (Treur 2020). Making use of the specific network characteristics offered by the self-modeling network structure format, a large number of options for different types of adaptation of mental models and different types of control over adaptation of mental models were obtained and structured. Many but not all of these options were illustrated by a several realistic examples that were already formalized by self-modeling networks. The options that were not illustrated here in a formalized computational sense, provide interesting options for a future research agenda.

References

Abraham, W.C., Bear, M.F.: Metaplasticity: the plasticity of synaptic plasticity. Trends Neurosci. **19**(4), 126–130 (1996)

Bartholomew, K., Horowitz, L.M.: Attachment styles among young adults: a test of a four-category model. J. Pers. Soc. Psychol. **61**(2), 226–244 (1991)

Bhalwankar, R., Treur, J.: Modeling the development of internal mental models by an adaptive network model. In: Proceedings of the 11th Annual International Conference on Brain-Inspired Cognitive Architectures for AI, BICA*AI'20. Procedia Computer Science. Elsevier (2021a)

Bhalwankar, R., Treur, J.: A second-order adaptive network model for learner-controlled mental model learning processes. In: Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds.) COMPLEX NETWORKS 2020 2020. SCI, vol. 944, pp. 245–259. Springer, Cham (2021b). https://doi.org/10.1007/978-3-030-65351-4_20

Bhalwankar, R., Treur, J.: 'If only i would have done that…': a controlled adaptive network model for learning by counterfactual thinking. In: Maglogiannis, I., Macintyre, J., Iliadis, L. (eds.) Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021, Hersonissos, Crete, Greece, June 25–27, 2021, Proceedings, pp. 3–16. Springer, Cham (2021c). https://doi.org/10.1007/978-3-030-79150-6_1

Bhalwankar, R., Van Ments, L., Treur, J.: Context-sensitive control of adaptation: self-modeling networks for human mental processes using mental models. In: Proceedings of the 3rd International Conference on Machine Learning and Intelligent Systems, MLIS 2021. Frontiers in Artificial Intelligence and Applications. IOS Press (2021)

Brogden, W.J.: Sensory preconditioning of human subjects. J. Exp. Psychol. **37**, 527–539 (1947)

Burtscher, M.J., Manser, T.: Team mental models and their potential to improve teamwork and safety: a review and implications for future research in healthcare. Saf. Sci. **50**(5), 1344–1354 (2012). https://doi.org/10.1016/j.ssci.2011.12.033

Chandra, N., Barkai, E.: A non-synaptic mechanism of complex learning: modulation of intrinsic neuronal excitability. Neurobiol. Learn. Mem. **154**, 30–36 (2018)

Craik, K.J.W.: The Nature of Explanation. University Press, Cambridge (1943)

Darling-Hammond, L., Austin, K., Cheung, M., Martin, D.: Thinking about thinking: metacognition. In: The Learning Classroom: Theory into Practice, pp. 157–172. Stanford University School of Education (2008)

Debanne, D., Inglebert, Y., Russier, M.: Plasticity of intrinsic neuronal excitability. Curr. Opin. Neurobiol. **54**, 73–82 (2019)

Du Plooy, M.C.: Visualisation as a metacognitive strategy in learning multiplicative concepts: a design research intervention. Ph.D. Thesis, University of Pretoria, Department of Mathematics Education (2016). https://repository.up.ac.za/handle/2263/51258

Hall, G.: Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. Anim. Learn. Behav. **24**(3), 233–255 (1996). https://doi.org/10.3758/BF03198973

Hebb, D.O.: The Organization of Behavior: A Neuropsychological Theory. Wiley (1949)

Hermans, A., Muhammad, S., Treur, J.: A second-order adaptive network model for attachment theory. In: Proceedings of the 21th International Conference on Computational Science, ICCS 2021. LNCS, vol. 12744, pp 462–475. Springer, Heidelberg (2021)

Hogan, K.E., Pressley, M.E.: Scaffolding Student Learning: Instructional Approaches and Issues. Brookline Books (1997)

Holland, J.H.: Adaptation in Natural and Artificial Systems. Ann Arbor, MI, University of Michigan Press (1975)

Hurley, S.: The shared circuits model (SCM): how control, mirroring, and simulation can enable imitation, deliberation, and mindreading. Behav. Brain Sci. **31**(1), 1–22 (2008)

Koedinger, K.R., Terao, A.: A cognitive task analysis of using pictures to support pre-algebraic reasoning. In: Gray, W.D., Schunn, C.D. (eds.) Proceedings of the 24th Annual Conference of the Cognitive Science Society, CogSci 2002, pp. 542–547. Lawrence Erlbaum Associates, Mahwah, NJ (2002)

Magerl, W., Hansen, N., Treede, R.D., Klein, T.: The human pain system exhibits higher-order plasticity (metaplasticity). Neurobiol. Learn. Mem. **154**, 112–120 (2018)

Mahdavi, M.: An overview: metacognition in education. Int. J. Multidisc. Curr. Res. **2**, 529–535 (2014)

March, J.G.: Exploration and exploitation in organizational learning. Organ. Sci. **2**, 71–87 (1991)

McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**(1), 415–444 (2001)

Piaget, J.: Origins of Intelligence in the Child (La Naissance de l'intelligence chez l'enfant). Routledge & Kegan Paul, London (1936)

Pintrich, P.R.: The role of goal orientation in self-regulated learning. In: Boekaerts, M., Pintrich, P., Zeidner, M. (eds.) Handbook of Self-regulation Research and Applications, pp. 451–502. Academic Press, Orlando, FL (2000)

Robinson, B.L., Harper, N.S., McAlpine, D.: Meta-adaptation in the auditory midbrain under cortical influence. Nat. Commun. **7**, e13442 (2016)

Seel, N.M.: Mental models in learning situations. In: Advances in Psychology, vol. 138, pp. 85–107. North-Holland, Amsterdam (2006)

Sjöström, P.J., Rancz, E.A., Roth, A., Hausser, M.: Dendritic excitability and synaptic plasticity. Physiol Rev **88**, 769–840 (2008)

Treur, J.: Network-Oriented Modeling for Adaptive Networks: Designing Higher-Order Adaptive Biological, Mental and Social Network Models. Springer Nature, Switzerland (2020)

Treur, J.: An adaptive network model covering metacognition to control adaptation for multiple mental models. Cogn. Syst. Res. **67**, 18–27 (2021a)

Treur, J.: Controlled social network adaptation: subjective elements in an objective social world. In: Yang, X.-S., Sherratt, S., Dey, N., Joshi, A. (eds.) Proceedings of Sixth International Congress on Information and Communication Technology. LNNS, vol. 235, pp. 263–274. Springer, Singapore (2021b). https://doi.org/10.1007/978-981-16-2377-6_26

Treur, J.: Self-modeling networks using adaptive internal mental models for cognitive analysis and support processes. In: Benito, R.M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L.M., Sales-Pardo, M. (eds.) Complex Networks 2020 . SCI, vol. 944, pp. 260–274. Springer, Cham (2021c). https://doi.org/10.1007/978-3-030-65351-4_21

Treur, J., Glas, G.: A multi-level cognitive architecture for self-referencing, self-awareness and self-interpretation. Cogn. Syst. Res. **68**, 125–142 (2021)

Treur, J., Van Ments, L. (eds.): Mental Models and their Dynamics, Adaptation, and Control: a Self-modeling Network Modeling Approach. Springer Nature, Heidelberg (2022, in press)

Van Hoeck, N., Watson, P.D., Barbey, A.K.: Cognitive neuroscience of human counterfactual reasoning. Front. Hum. Neurosci. **9**, 420 (2015)

van Ments, L., Treur, J.: A Higher-order adaptive network model to simulate development of and recovery from PTSD. In: Paszynski, M., Kranzlmüller, D., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds.) ICCS 2021. LNCS, vol. 12743, pp. 154–166. Springer, Cham (2021a). https://doi.org/10.1007/978-3-030-77964-1_13

Van Ments, L., Treur, J.: Modeling adaptive cooperative and competitive metaphors as mental models for joint decision making. Cogn. Syst. Res. **69**, 67–82 (2021b)

Van Ments, L., Treur, J.: Reflections on dynamics, adaptation and control: a cognitive architecture for mental models. Cogn. Syst. Res. **70**, 1–9 (2021c)

van Ments, L., Treur, J., Klein, J., Roelofsma, P.: A second-order adaptive network model for shared mental models in hospital teamwork. In: Nguyen, N.T., Iliadis, L., Maglogiannis, I., Trawiński, B. (eds.) ICCCI 2021. LNCS (LNAI), vol. 12876, pp. 126–140. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88081-1_10

Van Ments, L., Treur, J., Roelofsma, P.H.M.P.: An adaptive network model for formation and use of a mental god-model and its effect on human empathy, Chap. 11. In: Treur, J., Van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation, and Control: A Self-modeling Network Modeling Approach. Springer Nature, Heidelberg (2022, in press)

Wilson, R.C., Geana, A., White, J.M., Ludvig, E., Cohen, J.D.: Humans use directed and random exploration to solve the explore–exploit dilemma. J. Exp. Psychol. Gen. **143**(6), 2074–2081 (2014)