Part I
Introduction

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_1

1. Dynamics, Adaptation and Control for Mental Models: A Cognitive Architecture

Laila van Ments[1] and Jan Treur[2] ✉
(1) AutoLeadStar, Jerusalem, Israel
(2) Social AI Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
In this chapter, an overview of the wide variety of occurrences of mental models in the literature is discussed. They are classified according to two dimensions obtaining four categories of mental models: static-dynamic and world-mental, where static refers to mental models for static world states or for static mental states and dynamic refers to mental models for world processes or for mental processes. In addition, distinctions are made for what can be done by mental models: they can, for example, be (1) used for internal simulation, they can be (2) adapted, and these processes can be (3) controlled. This leads to a global three-level cognitive architecture covering these three ways of handling mental models. It is discussed that in this cognitive architecture reflection principles play an important role to define the interactions between the different levels.

Keywords Mental model – Cognitive architecture – Dynamics – Adaptation – Control

1.1 Introduction

Mental models are a kind of blueprints or pictures in the mind that can occur in various forms; e.g., Craik (1943), Evans (2006), Furlough and Gillan (2018), Gentner and Stevens (1983), Halford (1993), Johnson-Laird (1983). One relatively simple example is that you perceive the world state in front of you and

after closing your eyes you still see a picture of this world state in your mind. Another, more dynamic example is that you perceive an impressive course of events in front of you and after closing your eyes you see a kind of movie replay in your mind that replays this course of events. Although the notions of 'picture' or 'movie' provide an intuitive way to imagine what a mental model can be, for the general case such notions should not be taken literally but more in a metaphorical sense. For example, in a wider sense you can imagine a situation that you have never seen. Humans often use some form of mental model, as a blueprint or manual to handle situations. Well-known examples are operating a device or machine or software system, but also how to handle somebody else who needs to be handled based on some special personal 'user manual'. Still other examples are standard patterns learnt to solve certain types of problems in the context of certain disciplines, as so often are learnt at school.

All these examples show the wide variety of possibilities for mental models, usually described as structures consisting of collections or *networks* of certain *relations* that can be of various types. In this chapter this variety will be discussed, analysed and structured in some more detail in such a way that a basis is obtained for a cognitive architecture to handle mental models.

1.2 Mental Models and What They Model

In this section, part of the extensive literature on mental models is discussed and a structured overview is made based on distinguishing whether they consider an external world or an internal mental world and whether they model a static situation or a dynamic process.

1.2.1 Mental Models as Small-Scale Models Within the Head

For the history of the mental models area, often Kenneth Craik is mentioned as a central person. In his book (Craik 1943) he describes a mental model as a *small-scale model* that is carried by an organism within its head as follows:

> 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)

He emphasizes that such internal models use certain types of *relation-structure* that makes the mental model work in a way similar to how the real

world works:

> By "relation-structure" I do not mean some obscure non-physical entity which attends the model, but the fact that it is a physical working model which works in the same way as the process it parallels…Thus, the model need not resemble the real object pictorially; Kelvins' tide-predictor, which consists of a number of pulleys on levers, does not resemble a tide in appearance, but it works in the same way in certain essential respects… (Craik 1943, p. 51).

This similarity is depicted in Fig. 1.1, where the relation $M \rightarrow M'$ within the mental model corresponds to a similar relation $W \rightarrow W'$ in the world; by this similarity, the mental model faithfully represents the world.

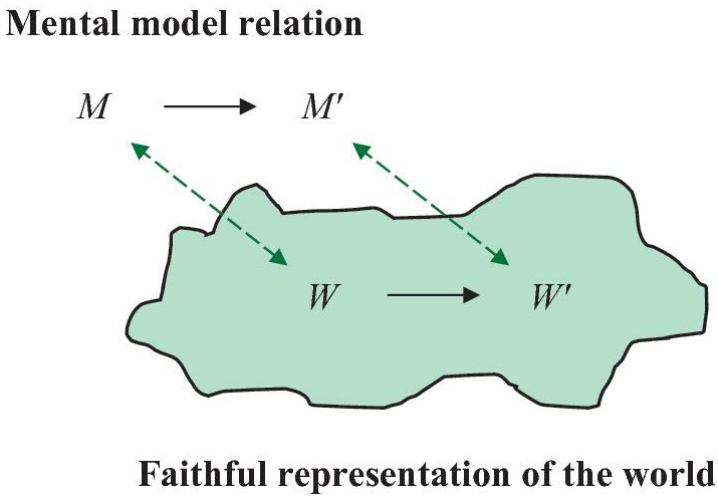


Fig. 1.1 A relation $M \rightarrow M'$ in a mental model and its correspondence to a similar relation $W \rightarrow W'$ in the world, thus providing a faithful representation of the world

In his book (Craik 1966) he emphasizes the benefits of being able to model and simulate processes from the world in the brain.

> "...the "modelling", by the brain, of the sequence of events whose consequence is sought, so that this model may predict the answer earlier than it occurs in the course of external nature. In other words, by the ability to model… the brain is able to outrun the physical processes which are too rapid for it and so can, on the average, forestall and anticipate the course of nature. (Craik 1966, p. 27).

For more discussion on Craik's work, see, for example Williams (2018a, b).

Other authors also have formulated what mental models are. For example, with an emphasis on causal relations, Shih and Alessi (1993, p. 157) explain that

> By a mental model we mean a person's understanding of the environment. It can represent different states of the problem and the causal relationships among states.

De Kleer and Brown (1983) describe it in the following way:

- The envisioning of the system, including a topological representation of the system components, the possible states of each of the components, and the structural relations between these components
- The running or execution of the causal model based on basic operational rules and on general scientific principles.

Moreover, after an extensive analysis, Doyle and Ford (1998) formulate the following definition, where the focus is on a dynamic system:

> A mental model of a dynamic system is a relatively enduring and accessible but limited internal conceptual representation of an external system whose structure maintains the perceived structure of that system.

All these descriptions strongly focus on how the world functions based on certain dynamic or temporal causal relations (sometimes called the *dynamic system view*) and that in a mental model similar relations are used to simulate a similar process. This idea of running a simulation inside the head is also called *internal simulation*; e.g., Damasio (1994), Goldman (2006), Hesslow (2002, 2012). For example, in Bhalwankar and Treur (2021a), the functioning of a car in interaction with its driver is internally simulated and in Van Ments and Treur (2021a) addressing PTSD, as a flashback experience a movie based on a mental model of a course of traumatic events is replayed in the brain; see also Treur and Van Ments (2022, Chap. 3), and Treur and Van Ments (2022, Chap. 5), respectively (this volume).

1.2.2 Mental Models for Individual Processes

In principle there are two ways in which a mental model can be considered to describe reasoning: to describe a reasoning state or to describe a reasoning process as a whole.

1.2.2.1 Describing Reasoning States by Mental Models

For example, Gentner and Stevens (1983), Johnson-Laird (1983), Halford (1993) put an emphasis on the use of mental models to reasoning states. Here *reasoning states* are the snapshots of a reasoning process at specific time points (sometimes also called information states or knowledge states). Each of these time-dependent reasoning states is conceptualized by a mental model or by a set

of mental models. Such mental models used to describe reasoning states have a slightly different appearance, compared to the mental models according to the perspective discussed above:

- In Sect. 1.2.1 dynamics is described within the mental model: the mental model represents the dynamics by relations defining a dynamical system
- In the current section the dynamics is not represented within a mental model: in contrast, a reasoning step is described as a transition step for mental models by which every step a new mental model is created.

So, based on this perspective where reasoning states are mental models, *reasoning steps* are considered transitions (maybe by standard generic inference rules, maybe based on other things) of one reasoning state to another one. For example, one reasoning state is described by a mental model

$$a$$
$$a \to b \qquad \text{where } \to \text{ denotes logical implication}$$

including the two items a and $a \to b$ both describing a static world situation and by a reasoning step (based on modus ponens in this case), this mental model is transformed into a new mental model

$$a$$
$$b$$
$$a \to b$$

including the three items a, b and $a \to b$ all describing the same static world situation. Thus reasoning steps are conceptualised as adaptations of the mental models representing these reasoning states, which can also have the form that two or more mental models are used (as antecedents) in combination. In a reasoning process as a whole, these transitions are executed in succession, resulting in sequences of reasoning states (also called *reasoning traces*). These are then conceptualised by sequences of mental models over time. See Johnson-Laird (2004) for more history of this perspective. This view on dynamics of reasoning also has been addressed within AI for different types of reasoning from a formal logical and computational perspective; for example, see Brazier et al. (1999), Gavrila and Treur (1994), Jonker and Treur (2002), Jonker and Treur (2003), Meyer and Treur (2001), Treur (1994).

1.2.2.2 Describing a Reasoning Process by One Mental Model

A different perspective on conceptualising reasoning processes can be obtained from the dynamical system view, more closely related to the view discussed in Sect. 1.2.1. From this view not a reasoning state like above, but instead a reasoning *process* is described as one mental model by temporal (or causal)

relations relating one reasoning state to another one. In other words, in this case a mental model describes the reasoning process by temporal relations defining a dynamical system, similar to how world processes can be modeled by a dynamical system mental model as discussed in Sect. 1.2.1. So, this time reasoning steps from one reasoning state to another one are not transitions between different mental models but are described by temporal relations within one fixed overall mental model. For example, such a mental model for a reasoning process can be described by relations of the form

$$a \rightarrow b \qquad (\text{where this time } \rightarrow \text{ is interpreted as a temporal casual relation})$$

expressing that within a reasoning process knowing *a* will causally affect knowing *b*, like how, for example, in philosophy of mind, in general mental states are assumed to causally affect each other over time; e.g., Kim (1996). This view on reasoning provides a description of reasoning like any other mental process as a dynamical system as described in Sect. 1.2.1. This perspective on reasoning is considered within literature in AI such as Engelfriet and Treur (1994), Engelfriet and Treur (1995), Meyer and Treur (2001), where a 'temporal theory of reasoning' plays the role of a mental model according to a dynamical system view.

So, in summary, in principle there are two different ways to model reasoning by mental models: (1) mental models model the reasoning states (as snapshots in a reasoning process), and reasoning steps are adaptations or transitions of these mental models over time, and (2) mental models model the dynamics of the reasoning process by modeling reasoning steps as temporal relations within one overall mental model for the entire reasoning process.

1.2.2.3 Mental Models Used to Model Cognitive Metaphors

Cognitive metaphors can also be considered a type of mental model (Cardillo et al. 2012; Carroll and Thomas 1982; Kuang 2003; Leary 1994; Ponterotto 2000; Romero and Soria 2005). Cognitive metaphors are a way to explain our conceptualization and mapping of new concepts on existing knowledge, and how we communicate this to others (Lakoff and Johnson 2003).

Or, in other words, one mental domain is understood in terms of a mental model of another phenomenon. Imagine encountering a novel animal: we will immediately compare its behaviors and looks to the bank of animals that exists in our brain, and try to map an understanding of this new animal, based on the knowledge we already have in our brain.

Another way to explain cognitive metaphors is as analogy making a mapping between a source and a target inside the brain (Gentner 1983; Gentner and Stevens 1983; Vosniadou and Ortony 1989), based on features the source and domain have in common. For example, we often hear the catchphrase 'Love is a

Journey'. Of course, literally, love is not a journey, but rather an abstract concept. However, because of the complexity of love as a concept, we use more concrete concepts, like a journey, to understand and communicate our understanding of love. By using the journey metaphor, we unconsciously map concepts like roadblocks and the fact that a journey is something to embark on onto our concept of love, and thus come closer to a (mutual) understanding. Lakoff (1993) stresses that metaphors are an essential mechanism that is systematically mapped in our brain for humans to understand the world, and be able to think and reason, without us even noticing. Furthermore, bodily changes can unconsciously affect our metaphorical thoughts, see Barsalou (2008), Landau et al. (2010), Williams et al. (2009). Even more so, our mental models can be influenced by the metaphors we use, as constant repetition of particular metaphors will lead to our unconscious acceptance of that particular metaphor as a normal way of seeing that situation El Refaie (2003). Thereby, a metaphor subconsciously constructs how we perceive situations; see Barsalou (2008), Landau et al. (2010), Williams et al. (2009). Several studies have shown that our actions are subconsciously influenced by the automated activation of motives (Bargh et al. 2001; Bargh and Morsella 2008). Therefore, through this route cognitive metaphors also affect the way humans make decisions. As an example, in (Van Ments and Treur 2021b) metaphors for cooperative and competitive joint decision making are modeled as a second-order adaptive mental model; see also (Treur and Van Ments 2022, Chap. 10) (this volume).

1.2.3 Mental Models in Social Processes

In this section the focus is on mental models used in a social context. These can concern mental models for bonding and attachment in dyadic relationships or mental models for groups, teams or organisations. A well-known social type of mental model occurs when one has some 'image' of the mental state of another person, or of oneself. If the dynamics of mental processes are also considered, one can, for example, have a mental model of how your partner will get angry or disappointed after you undertake some specific action. Also more in general in social life, humans often use some mental model to understand each other and interact in an adequate manner based on that mental model, for example, to get something done. And the same even applies to having a user manual for handling oneself. In this section, in particular mental models for attachment in dyadic relationships are briefly discussed, and mental God-models, mental models for bonding by homophily, and team mental models.

1.2.3.1 Mental Models for Attachment

The way an individual forms relationships with others can be explained by the *Attachment Theory*, constructed by Ainsworth and Bowlby. This theory is based

on its predecessor, the 'Security Theory', developed by William Blatz and Mary Salter Ainsworth; e.g., Blatz (1966), Salter (1940), Salter Ainsworth (2010), Salter Ainsworth and Bowlby (1965). The attachment theory explains how a child develops a set of emotions, memories, thoughts, expectations, behaviours and beliefs about itself and others, based on its early experiences with its primary caregiver. This set is called the 'internal working model of social relationships', which continues to change with age and experience (Mercer 2006). More specifically, this 'model of self' and 'model of other' that the child initially develops, is based on experiences with the primary caregiver and their behaviour (Bretherton 1992). Using their internal 'model of other' children can predict the primary caregiver's behaviour, and using their internal 'model of self', they can plan their own behaviour accordingly (Bretherton 1992), and the same happens later in life in interaction with significant others.

In Hermans et al. (2021) a second-order adaptive network model is presented for development of mental models of self and others according to Attachment Theory; see also (Treur and Van Ments 2022, Chap. 12) (this volume).

1.2.3.2 Mental God-Models

Another interesting place where we can find place we can find is double mental models is a person's relationship with God. When a person prays, the same brain regions that are used for interactions with other people become activated, enabling a person to generate an internal representation of 'the other', in this case the image they have of God. This allows people to form a real, meaningful relationship with God, and to construct a mental model of an image of God (Schjoedt et al. 2009). This mental God-model that an individual has of God, and how this image has impact on the individual, can involve many aspects. For example, the attachment style discussed in the previous section can be studied in combination with a person's God-model, and how these two influence each other (Granqvist and Kirkpatrick 2008). The relationship and mental image of God can also be explained from a mental model or mentalizing perspective, as introduced by Schaap-Jonker and Corveleyn (2014). Mentalizing is the capacity of thinking about thinking and feeling. It provides awareness that one's own and others' behaviour is driven by mental states, and gives the ability to selectively activate internal states that fit the individual's particular. Mentalizing also involves a process of internal simulation, where an individual internally simulates mind states to predict effects in the external world or other persons. In other words, a mental model is an interesting way to describe an individual's relationship with God.

In Van Ments et al. (2022) an adaptive network model for developing and using a mental God-model is described.

1.2.3.3 Mental Models for Bonding Based on Homophily

Social networks often are adaptive, for example based on a bonding by a homophily principle for the adaptation of the weights of the network connections between persons over time. A bonding by homophily adaptation principle expresses how 'being alike' strengthens the connection between two persons, also explained as 'birds of a feather flock together'; e.g., McPherson et al. (2001). Usually, in literature such adaptation processes are considered without taking into account subjective elements for the persons involved. For example, do the persons themselves actually know in how far they are alike? Or are they just will-less victims of objective social laws independent of what they know or what they want? Such subjective aspects are often lacking in (computational) research on bonding by homophily, as usually these processes are addressed exclusively from the perspective of an assumed objective social world. However, a more realistic bonding by homophily principle can be obtained if the bonding is not assumed to be based on an objective form of homophily but on the mental models both persons have of each other. If two persons both have a mental model of themselves and the other that show that they are alike, then that will clearly affect their bonding, even if these mental models are not correct and, for example, based on fake information. This subjective mental model based perspective on bonding by homophily is addressed in Treur (2021b); see also Treur and Van Ments (2022, Chap. 13), (this volume), which also includes an example scenario where one of the persons on purpose fakes incorrect personal characteristics or properties in order to make bonding happen.

1.2.3.4 Team Mental Models

A team mental model is based on the assumption that high performing teams need to have team members that are on the same page in order to perform complex tasks well; e.g., Burtscher and Manser (2012), Langan-Fox et al. (2000), Mohammed et al. (2010). This requires team members to have a shared understanding of the relevant elements to perform a specific task. A team mental model is an emergent team level concept which is generated by each team member's cognition up to the level that it becomes a shared mental model: so, the origin and basis of a team mental model is formed by the individual team members. More specifically, the team mental model itself is an emerging collective phenomenon which is created bottom-up from each team member's cognition in a dynamic manner (DeChurch and Mesmer-Magnus 2010a, b). The main functions of team mental models are improved planning, coordination and alignment (Nini 2019). Two types of team mental models are distinguished (Mohammed et al. 2010):

- task-related team mental models

- team-related team mental models.

The first type provides a team's cognitive representation of task-related elements such as goals and subtasks, subtask dependencies, subtask durations, milestones, and resources required for task coordination. The second type covers the team's mental model for the knowledge, skills, competencies and relationships of team members.

In Van Ments et al. (2021) an example of an adaptive network model for handling a team mental model in a medical context is presented; see also Treur and Van Ments (2022, Chap 14) (this volume).

1.2.4 A Mental Models Overview According to Mental Versus World and Static Versus Dynamic

In the above Sects. 1.2.1–1.2.3, mental models have been described as consisting of a collection or network of relations. In some cases these relations describe static relationships for a world situation or state (such as 'Joe is taller than Kamala') or for a mental state (such as 'Joe does not believe in complot theories'). In other cases, these relations describe temporal or causal relationships according to a dynamic system view of a world process (e.g., 'human action causes climate change') or a reasoning process (e.g., 'because I believe I have no time left, I now decide to do this action'). In all such examples, that can be represented by a mental model, two dimensions of variation can be recognized. The first dimension is the dimension *static-dynamic*, where static refers to representing static situations, and dynamic to representing a process. The second dimension is the dimension *world-mental* where world refers to the external world and mental to mental states or processes. Distinctions according to these dimensions have been used in Table 1.1 to get a structured overview of the options.

Table 1.1 The variety of mental models structured for what is modeled according to state vs process and world versus mental; this provides a summary of the concepts discussed in more detail in the text of Sect. 1.2

State versus process	World versus mental	Example mental models
Process description	World process	• A mental model of a dynamical system for world dynamics • A mental model of a how the water level changes with tide • A mental model of how the climate of the earth changes due to human action • A step-by-step description of a route to follow to get from A to B in a city; e.g., 'when you reach the cinema on your right hand, turn left and get into that street to the supermarket'

State versus process	World versus mental	Example mental models
	Mental process	• A mental model of how your partner will get angry or disappointed after you undertake a specific action • A mental model of how you yourself will get angry or disappointed after your partner undertakes a specific action • A step-by-step algorithm to calculate the area of a rectangle or a long division
State description	World state	• A mental model of a city in the form of a map; e.g., 'the supermarket is in the street opposite the cinema' • A mental model of the current climate in different regions • A mental model of a rectangle
	Mental state	• A mental model of beliefs someone else has on the world • A mental model of desires or goals someone else has • A mental model of the emotions someone else has • A mental model of the knowledge and skills someone else has • A mental model of any of the above for yourself instead of 'someone else'

Note that this table is not the end of the story, as several important aspects that occur in relation to mental models are not covered yet. As mental models are usually described as networks of certain types of relations, one characteristic of mental models that also varies is which types of relation are used exactly. Causal relations are often used, especially from a dynamic system view, but also other types of relations often occur in mental models. In addition, also relations of higher-order, as used among others, in analogical reasoning have not been distinguished yet. Moreover, the adaptation of mental models as takes place in learning or development still has to be addressed, and the same holds for the control over such adaptation. These topics will be addressed in next sections.

1.3 Learning and Development of Mental Models

Within educational science, mental models are often considered an important vehicle for learning; for just a few of the many contributions, see Benbassat (2014), Buckley (2000), Doll et al. (2012), Du Plooy (2016), Greca and Moreira (2000), Halloun (1996), Hurley (2008), Koedinger and Terao (2002), Larbi and Mavis (2016), Mayer (1989), Seel (2006), Skemp (1971), Van Gog et al. (2009), Yi and Davis (2003). The focus in this area is usually on how mental models can be formed (learnt) and adapted over time. In this section this perspective from educational science is discussed in some detail.

1.3.1 Learning and Development as Adaptation of Mental Models

Within educational science, sometimes the term *model-based learning* is used for learning described as constructing coherent mental models; for example, Buckley (2000) formulates this as:

> Model-based learning is a dynamic, recursive process of learning by building mental models. It incorporates the formation, testing, and subsequent reinforcement, revision, or rejection of mental models of some phenomenon.

However, note that in most cases that mental models are considered for learning, the term model-based learning is not explicitly used. This view on learning was also described by Piaget. Although he did not use the term mental model, the ideas he put forward do apply to mental models. Within the literature also the term schema or schemata is often used; this concept has no sharp boundary with the concept mental model and both concepts have much in common. Following the ideas of Piaget (1936, 1954) , formation and adaptation of mental models during learning or development can occur in two forms: by *assimilation* (extension or refinement of a mental model) or by *accommodation* (revision of a mental model). As an example, suppose that a mental model includes the relation.

$$\text{need something} \rightarrow \text{go to shopping area}$$

By assimilation, this can be refined into a mental model including the following relations:

$$\text{need something} \rightarrow \text{go to shopping area}$$
$$\text{need book} \rightarrow \text{need something}$$
$$\text{need book} \,\&\, \text{in shopping area} \rightarrow \text{look for book shop}$$

This is a refinement and not a revision, as the previous relation still applies. In contrast, accommodation takes place, for example, when due to a lockdown the shops are closed for a long time. Then the mental model including.

$$\text{need something} \rightarrow \text{go to shopping area}$$

can be revised into a mental model including.

$$\text{need something} \rightarrow \text{go to web shop}$$

This is indeed a revision and not a refinement as the previous relation does not apply anymore. Such types of examples illustrate how mental models can change over time due to learning or development, as also described by the quote above from Buckley (2000). Next, some elements of learning processes are

addressed in more detail and the importance of control over the learning is discussed.

1.3.2 Learning of Mental Models by Observation and by Instruction

Observational learning indicates when observation is important for the learning or development of a mental model. This can be observation of others but also observation of oneself while 'learning by doing' or 'learning by discovery'. Learners may see someone perform a t type of behavior and then start to imitate it; e.g., Benbassat (2014), Yi and Davis (2003). This is often used to make others learn a specific motor task. A mechanism based on mirror neurons underly the ability to learn by observing and imitating others; e.g., Hurley (2008), Rizzolatti and Craighero (2004) Van Gog et al. (2009). An example of an adaptive network model for learning by observation a mental model of how a car works and how to drive it can be found in Bhalwankar and Treur (2021a); see also Treur and Van Ments (2022, Chap. 3) (this volume). Another example showing how a mental model is learned by counterfactual thinking and observation can be found in Bhalwankar and Treur (2021c); see also Treur and Van Ments (2022, Chap. 6) (this volume).

Instructional learning describes how information provided by an expert instructor can be an important source for the learning. Only learning based on observation often may lead to processes of trial and error; e.g., Seel (2006). Instructions from an expert are a useful addition to develop mental models in an effective manner. A format of scaffolded model-based learning in which many supporting actions such as prompts, questions, hints, stories, conceptual models, visualizations are performed, facilitates a learner's progress; e.g., Hogan and Pressley (1997). An example of an adaptive network model for learning by instruction a mental model of how a car works and how to drive it can be found in Bhalwankar and Treur (2021a); see also Treur and Van Ments (2022, Chap. 3) (this volume).

1.3.3 Control for Learning of Mental Models Based on Metacognition

To handle mental models and in particular the learning of them, *control* is important; e.g., Gibbons and Gray (2002) claim that instructions are most effective for learning processes when the learner controls them. The already mentioned scaffolded model-based learning format in the previous section supports this (Hogan and Pressley 1997). As another example, Kozma (1991) claims that persons actively pick external sources for mental model learning. So, the learner's initiatives for instruction and information acquisition are

important for mental model learning. The learner has (to be able) to be proactive and in control of the learning. As yet another example, Meela, and Yuenyong (2019) have shown that Model-Based Inquiry (MBI) can support a student's mental model formation in scientific learning; see also Neilson et al. (2010).

An example of an adaptive network model for controlled learning of a mental model of how a car works and how to drive it can be found in Bhalwankar and Treur (2021b); see also Treur and Van Ments (2022, Chap. 9).

Metacognition is described in Darling-Hammond et al. (2008), Shannon (2008), Mahdavi (2014), Flavell (1979), Koriat (2007), Pintrich (2000) as cognition about cognition. More specifically, Koriat (2007) presents it as what people know about their own cognitive processes and how they put that knowledge to use in regulating their cognitive processing and behavior. Sometimes the term self-regulation and self-regulated learning are used. In Pintrich (2000), this is formulated as an active, constructive process whereby learners set goals for their learning and then monitor, regulate, and control their cognition, motivation, and behavior, guided by these goals.

Also in learning complex tasks using mental models, control is a crucial element; see Treur (2021c) for an example network model for this; see also Treur and Van Ments (2022, Chap. 7) (this volume).

In learning, often different mental models play a role; e.g., Gentner and Stevens (1983), Greca and Moreira (2000), Skemp (1971), Seel (2006). An example can be the learning of subtracting numbers. The learner can use a more visual model, drawing out the numbers on a line, or a more abstract model, using formulas to represent the subtraction e.g., Bruner (1966), Du Plooy (2016). Here, metacognition plays an important role for the decisions about when to *switch* from one mental model to another one. In Treur (2021a) more can be found on this case, particularly for learning arithmetic or algebraic skills in primary or secondary schools supported by visualisation; see also Treur and Van Ments (2022, Chap. 4) (this volume).

1.4 A Cognitive Architecture for Mental Models

In this section several aspects of mental models are discussed that are important to obtain a cognitive architecture to handle mental models. In particular, the following aspects are addressed:

- higher-order relations in mental models
- adaptation of mental models
- control of adaptation of mental models.

Finally, it will be pointed out how an overall cognitive architecture can be designed covering these aspects.

1.4.1 Higher-Order Relations

Higher-order relations are relations between relations. In Fig. 1.2 an example is depicted of a first-order relation R and a second-order relation T. In this example, this second-order relation T expresses that the first-order relation R is transitive. Below the dashed purple line, a first-order mental model is depicted based on relation R. Above this dashed purple line a second-order mental model is depicted based on transitivity relation T.

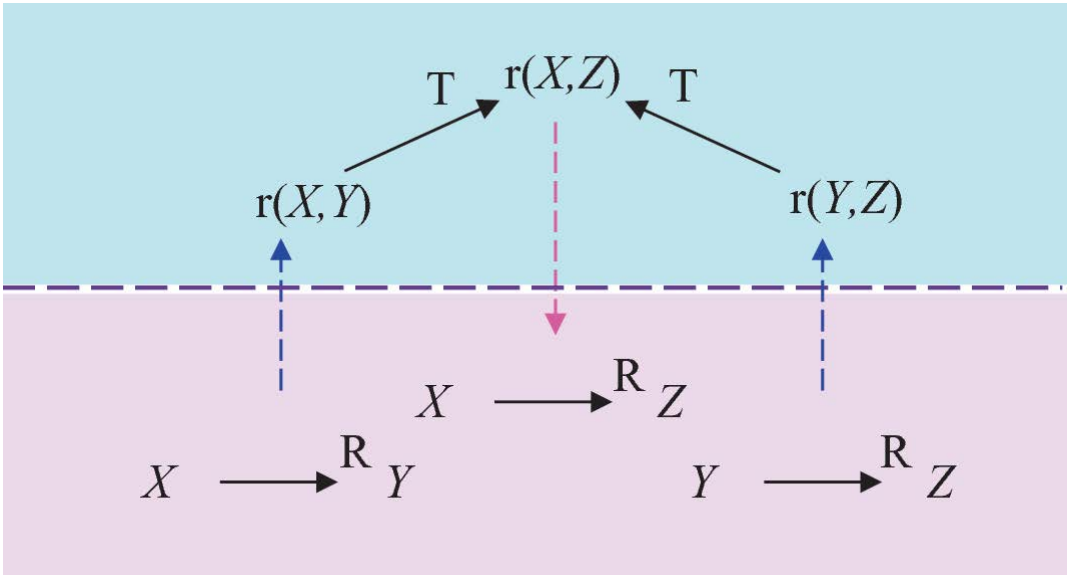


Fig. 1.2 Second-order relation *T* expressing transitivity of first-order relation *R*

In the first-order self-model, the relation R is a relation for objects *X*, *Y* and *Z*, where, for example, R denotes the relation 'is taller than'. Linguistically or logically, such a first-order relation can also be expressed as *X*:*Y* or $X:_{\mathrm{R}}Y$ or *X* R *Y* or R(*X*, *Y*). In the second-order mental model, the relation T is also between certain objects, but this time the objects are indicated by terms r(*X*,*Y*), r(*Y*,*Z*), and r(*X*,*Z*) which are names for the relation instances $X \xrightarrow{\mathrm{R}} Y$, $X \xrightarrow{\mathrm{R}} Y$, and $X \xrightarrow{\mathrm{R}} Z$, respectively, one level lower. These objects can be considered reifications of the relation instances represented in the first-order mental model: they are now represented by objects like r(*X*, *Y*) that refer to relational expression R(*X*, *Y*); e.g., see Galton (2006). This is similar to, for example, how Gödel used a representation of logical statements by natural numbers to obtain his famous incompleteness theorems for mathematical logic; e.g., see Hofstadter (1979), Nagel and Newman (1965), Smorynski (1977). The upward and downward relations between the two levels can be described by so-called *reflection principles*; see also Sect. 1.4.3 below and Treur (1991, 1994), Weyhrauch (1980).

Another example of a second-order relation in a slightly different notation is the relation $A{:}B{::}C{:}D$ where the symbol : denotes the first-order relation and the symbol :: denotes a second-order relation between the two first-order relational expressions $A{:}B$ and $C{:}D$. This is often used in experiments concerning analogical inference as also discussed in Sect. 1.4.2; e.g., Alfred et al. (2020), Holyoak and Monti (2020), Whitaker et al. (2018). For such a second-order relation $A{:}B{::}C{:}D$, a picture similar to Fig. 1.2 can be drawn. In principle, also third- and higher-order relations may be possible; the use of third-order relations for control is discussed in Sect. 1.4.3.

The above shows that in addition to the distinctions made in Table 1.1, also a distinction between mental models according to the orders of the relations they use can be made, where in one mental process multiple mental models of different orders may be used in an integrative manner.

1.4.2 What Exactly Do Mental Models Do?

Next, distinctions are made for what mental models actually do. In different sections, different types of processes were encountered that in one way or the other relate to mental models. The following overview of these processes can be made.

- **Simulation: Mental Models Simulate**
  As discussed in Sects. 1.2 and 1.4, mental models are often used for a form of inferencing or internal or mental simulation to relate known facts to unknown facts about world or mental states or processes. This occurs in many forms, varying from prediction, visualisation in sport, flashback movies in PTSD, dreaming, reasoning and many more cases.
- **Adaptation: Mental Models Adapt**
  Mental models often are adapted; they can be formed or learned and they can be revised, as Piaget (1936, 1954) already pointed out. This has been discussed in some detail in Sect. 1.3, thereby addressing observational learning and instructional learning in particular.
- **Control: Mental Models Respond to Control**
  Using mental models and adapting them is in principle done in a coordinated manner by some form of control by a form of metacognition. This also has been discussed in some detail in Sect. 1.3 in particular for the timing of observational learning and instructional learning.

This shows that in addition to the distinctions made in Table 1.1 and in Sect. 1.4.1, also distinctions have to be made between what mental models actually do, where in one mental process often multiple mental models of different levels will be used in interaction with each other. In Sect. 1.4.3, it is pointed out how a cognitive architecture for this may be obtained.

1.4.3 A Cognitive Architecture for Handling Mental Models

Based on the different processes in which mental models are used as summarised in Sects. 1.4.1 and 1.4.2, it can be assumed that a cognitive architecture for handling mental models has to cover the following three types of processes in an integrated manner (see also Fig. 1.3):

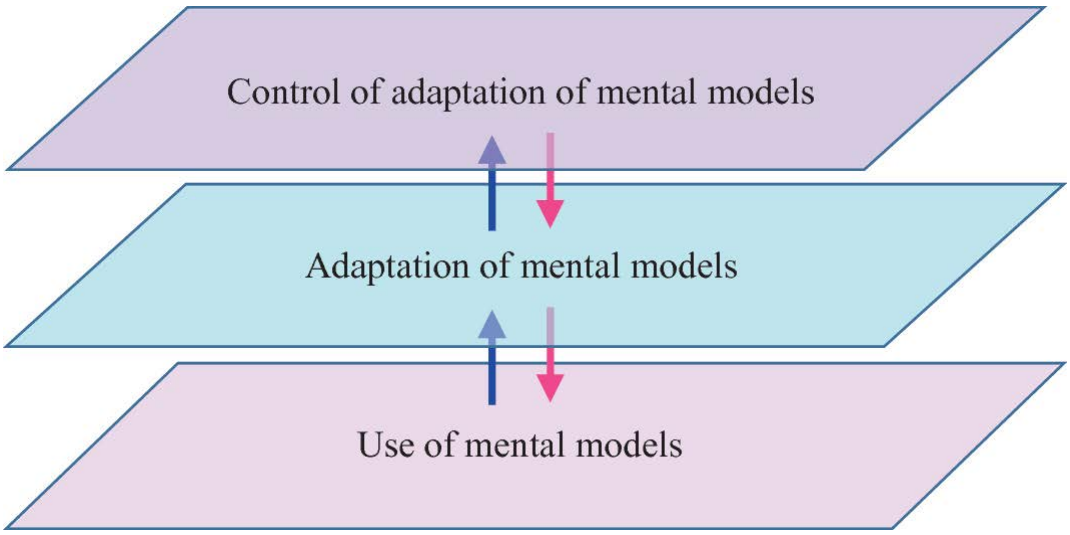


Fig. 1.3 Cognitive architecture for mental model handling with three levels of mental processing for mental models where each next level is modeled by relations one order higher than at the level below it

Level 1: Use (Base Level)
This level covers mental models described by relations that can be used to generate internal simulation.

Level 2: Adaptivity (First-Order Adaptation Level)
This level covers adaptation of Level 1 mental models by learning, revision, or other change; this can be described by a mental model using relations for changing the relations of the mental models at Level 1. In principle, this will involve a mental model with relations of one order higher than the relations used at Level 1.

Level 3: Control (Second-Order Adaptation Level)
This level covers control of adaptation processes described by a mental model using relations for changing the relations used at Level 2 for change of the Level 1 mental models. In principle, this will involve relations of one order higher than the relations used at Level 2, and two orders higher than the relations used at Level 1.

Here the second and third level are higher-order levels (involving higher-order relations; see Sect. 1.4.1) compared to the first level. This architecture was inspired by literature on metalevel architectures and reflection such as Bowen

and Kowalski (1982), Bowen (1985), Galton (2006), Sterling and Beer (1989), Treur (1991), Treur (1994), Weyhrauch (1980). To illustrate the levels in Fig. 1.3 and their relations by an abstract mini-example, assume at the three levels 1 to 3 relations R, S and T (denoted by $\xrightarrow{R}$, $\xrightarrow{R}$, $\xrightarrow{R}$, respectively) are used as shown in Table 1.2 (columns 2–4) and Fig. 1.4; here *V, W, X* (column 5) model some contextual or situational factors. These relations may be causal relations, but they can also be of any other type of relation. An important notion to describe the interaction between the different levels of such an architecture is the notion of *reflection principle* (Treur 1991, 1994; Weyhrauch 1980); this type of principle (see also column 6 in Table 1.2) will also be explained below by the mini-example

Table 1.2 Overview of the mini-example for the three levels

Level	Relations	Relation instances	Object terms	Context	Reflection principles
3	T	$V \xrightarrow{R} s(W, r(X, Y))$	$V, s(W, r(X, Y))$	V	$s(W, r(X, Y))$ $\uparrow\downarrow$ $r(X, Y)$ $W \xrightarrow{S} r(X, Y)$ $\uparrow\downarrow$ $X \xrightarrow{R} Y$
2	S	$W \xrightarrow{S} r(X, Y)$	$W, r(X, Y)$	W	
1	R	$X \xrightarrow{R} Y$	X, Y	X	

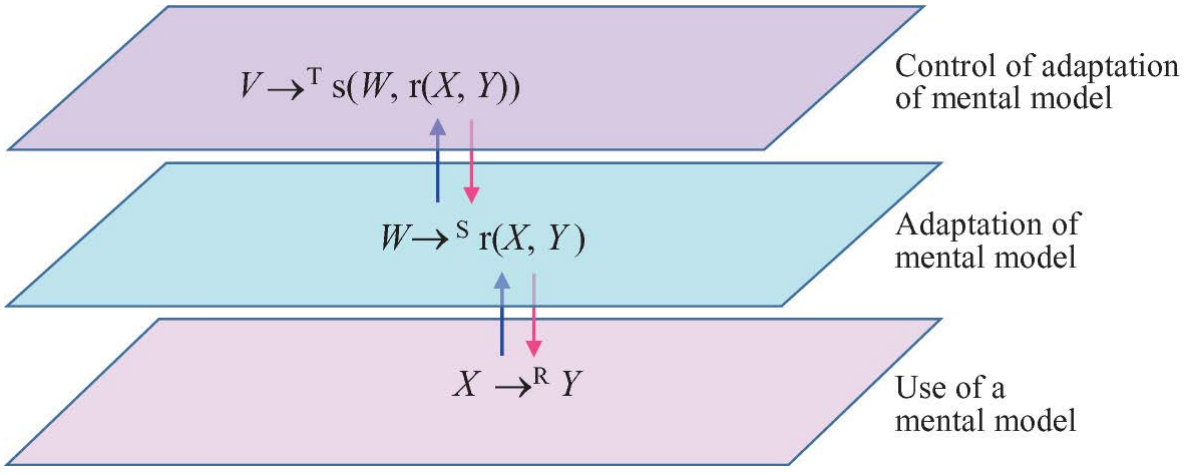


Fig. 1.4 The mini-example within the cognitive architecture

The explanation of this mini-example is as follows. At the base level the mental model includes an instance of relation R from *X* to *Y*, represented as

$$X \xrightarrow{R} Y$$

This relation R is usually called a first-order relation. By an upward reflection principle from level 1 to level 2, at the second level (for adaptation) this R-relation instance relates to an object denoted by the term

$$r(X, Y)$$

referring to relation $X \xrightarrow{\mathrm{R}} Y$; so, r(*X*, *Y*) is a name to refer to relation instance $X \xrightarrow{\mathrm{R}} Y$(alternatively, sometimes the notation '$X \xrightarrow{\mathrm{R}} Y$' is used for such a name). For this object at level 2, in turn an instance of relation S applies that relates the object r(*X*, *Y*) to context factor *W*:

$$W \xrightarrow{\mathrm{S}} \mathrm{r}(X, Y)$$

This relation S is usually called a second-order relation. By a downward reflection principle from level 2 to level 1, this makes first-order relation R adaptive, as via the relation $W \xrightarrow{\mathrm{S}} \mathrm{r}(X, Y)$ the object $\mathrm{r}(X, Y)$ representing $X \xrightarrow{\mathrm{R}} Y$ depends on circumstances modeled by context factor *W* and by the downward reflection principle, this affects the relation instance $X \xrightarrow{\mathrm{R}} Y$ at level 1 accordingly. Note that for this cognitive architecture, this is called first-order *adaptation*, as it concerns adaptation of the first-order relation. But note that the term used in the literature for the relation S involved is *second-order relation*.

However, also the second-order relation S is adaptive, because similarly by an upward reflection principle from level 2 to level 3 it relates to an object denoted by the term

$$\mathrm{s}(W, \mathrm{r}(X, Y))$$

at level 3 referring to relation $W \xrightarrow{\mathrm{S}} \mathrm{r}(X, Y)$, and this object also depends on circumstances (modeled by context factor *V*), as at level 3 a third-order relation T is applied:

$$V \xrightarrow{\mathrm{T}} \mathrm{s}(W, \mathrm{r}(X, Y))$$

Therefore, $\mathrm{s}(W, \mathrm{r}(X, Y))$ depends on context factor *V* and by a downward reflection principle from level 3 to level 2, this affects S-relation instance $W \xrightarrow{\mathrm{S}} \mathrm{r}(X, Y)$ at level 2 accordingly. As second-order relation S models the first-order adaptation of first-order relation R, by this control over the first-order adaptation can be exerted. In summary, second-order relation S models adaptation of first-order relation R using context factor *W*, whereas third-order relation T models control of this adaptation using context factor *V*. Note that for this cognitive architecture, this is called *second-order adaptation*, as it concerns adaptation of the second-order relation. But the term used in the literature for the relation T involved is *third-order relation*.

This simple example illustrates how the adaptation of a mental model and its control can be modeled, and it points out how reflection principles can connect the levels and enable the transfer between the levels.

This structure of three levels for handling mental models can be used in conjunction with the structure of Table 1.1 in Sect. 1.2 to obtain an overview of the many possible occurrences and uses of mental models. Note that due to the relationships between the different levels explained above where objects at each higher level refer to relations at the next lower level, the higher levels can be interpreted as *self-models* of part of the architecture itself, namely self-models of the part at the next lower level. In this sense it can be considered a *self-modeling architecture*. In Treur and Van Ments (2022, Chap. 2) (this volume), using the notion of multi-level *self-modeling network* (also called reified network), it is described in more detail how this cognitive architecture with three description levels indeed can be modeled based on a three-level self-modeling network model. Moreover, in Treur and Van Ments (2022, Chap. 21) (this volume) a more in depth analysis is presented on what the self-modeling network format can offer for modeling the cognitive architecture introduced here and its applications.

1.5 Discussion

In this chapter, an overview of the wide variety of occurrences of mental models in the literature was discussed. They were classified according to two dimensions obtaining four categories of them: static-dynamic and world-mental, where static refers to mental models for static world states or for static mental states and dynamic refers to mental models for world processes or for mental processes. In addition, distinctions were made for what can be done by mental models: they can, for example, be (1) used for internal simulation, they can be (2) adapted, and these processes can be (3) controlled. This has led to a global three-level cognitive architecture covering these three ways of handling mental models. It has been pointed out that in this cognitive architecture reflection principles play an important role to define the interactions between the different levels. In Treur and Van Ments (2022, Chap. 2), the notion of self-modeling network is used to work this architecture out in more detail based on the self-modeling network modeling approach described in Treur (2020). For this modeling approach, further details on design using the modeling environment can be found in Treur and Van Ments (2022, Chap. 17) (this volume), on verification by analysis of stationary points and analysis in Treur and Van Ments (2022, Chap. 18), validation using parameter tuning in Treur and Van Ments (2022, Chap. 19), and the scope of applicability in Treur and Van Ments (2022, Chap. 20). Note that in many cases the three-level cognitive architecture described in the current chapter is sufficient. However, sometimes a model with more than three levels fits better, as, for example, shown in Treur and Van Ments (2022, Chap. 8).

Some more philosophically focused background for mental models and their modeling can be found in Treur (2021d) about neural correlates for mental models and (Treur 2021e) about the emerging informational content of mental models; see also Treur and Van Ments (2022, Chap. 15), and Treur and Van Ments (2022, Chap 16), respectively.

References

Alfred, K.L., Connolly, A.C., Cetron, J.S., Kraemer, D.J.M.: Mental models use common neural spatial structure for spatial and abstract content. Commun. Biol. **3**, 17 (2020)

Barsalou, L.W.: Grounded cognition. Annu. Rev. Psychol. **59**(1), 617–645 (2008)

Bargh, J.A., Gollwitzer, P.M., Lee-Chai, A., Barndollar, K., Trötschel, R.: The automated will: nonconscious activation and pursuit of behavioral goals. J. Pers. Soc. Psychol. **81**(6), 1014–1027 (2001)

Bargh, J.A., Morsella, E.: The Unconscious mind. Perspect. Psychol. Sci. **3**(1), 73–79 (2008)

Benbassat, J.: Role modeling in medical education: the importance of a reflective imitation. Acad. Med. **89**(4), 550–554 (2014)

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, vol. 190, issue 4, pp. 90–101 (2021a)

Bhalwankar, R., Treur, J.: A second-order adaptive network model for learner-controlled mental model learning processes. In: Proceedings of the 9th International Conference on Complex Networks and their Applications, vol. 2. Studies in Computational Intelligence, vol. 944, pp. 245–259. Springer, Switzerland AG (2021b)

Bhalwankar, R., Treur, J.: If only i would have done that...: A controlled adaptive network model for learning by counterfactual thinking. In: Proceedings of the 17th International Conference on Artificial Intelligence Applications and Innovations, AIAI'21. Advances in Information and Communication Technology, vol. 627, pp. 3–16. Springer (2021c)

Blatz, W.E.: Human Security: Some Reflections. University of Toronto Press, Toronto, Canada (1966)

Bowen, K.A., Kowalski, R.: Amalgamating language and meta-language in logic programming. In: Clark, K., Tarnlund, S. (eds.) Logic Programming, pp. 153–172. Academic Press, New York (1982)

Bowen, K.A.: Meta-level programming and knowledge representation. N. Gener. Comput. **3**, 359–383 (1985)

Brazier, F.M.T., Treur, J., Wijngaards, N.J.E., Willems, M.: Temporal semantics of compositional task models and problem solving methods. Data Knowl. Eng. **29**(1), 17–42 (1999)
[zbMATH]

Buckley, B.C.: Interactive multimedia and model-based learning in biology. Int. J. Sci. Educ. **22**(9), 895–935 (2000)

Bretherton, I.: The origins of attachment theory: John Bowlby and Mary Ainsworth. Dev. Psychol. **28**, 759–775 (1992)

Bruner, J.S.: Towards a Theory of Instruction. Harvard University, Cambridge, Mass (1966)

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. Safety Sci. **50**(5), 1344–1354 (2012). https://doi.org/10.1016/j.ssci.2011.12.033

Cardillo, E.R., Watson, C.E., Schmidt, G.L., Kranjec, A., Chatterjee, A.: From novel to familiar: tuning the brain for metaphors. Neuroimage **59**(4), 3212–3221 (2012)

Carroll, J.M., Thomas, J.C.: Metaphor and the cognitive representation of computing systems. IEEE Trans. Syst. Man Cybern. **12**(2), 107–116 (1982)

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

Craik, K.J.W.: In: Sherwood, S.L (ed.) The Nature of Psychology. Cambridge University Press, Cambridge (1966)

Damasio, A.R.: Descartes Error: Emotion, Reason and the Human Brain. Vintage Books, London (1994)

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)

DeChurch, L.A.; Mesmer-Magnus, J.R.: Measuring shared team mental models. A meta-analysis. Group Dyn.: Theory Res. Practice **14**(1), 1–14 (2010a). https://doi.org/10.1037/a0017455

DeChurch, L.A.; Mesmer-Magnus, J.R.: The cognitive underpinnings of effective teamwork. A meta-analysis. J. Appl. Psychol. **95**(1), 32–53 (2010b). https://doi.org/10.1037/a0017328

De Kleer, J., Brown, J.: Assumptions and ambiguities in mechanistic mental models. In: Gentner, D., Stevens, A. (eds.) Mental Models, pp. 155–190. Lawrence Erlbaum Associates, Hillsdale, NJ (1983)

Doll, B.B., Simon, D.A., Daw, N.D.: The ubiquity of model-based reinforcement learning. Curr. Opin. Neurobiol. **22**, 1075–1081 (2012)

Doyle, J.K., Ford, D.N.: Mental models concepts for system dynamics research. Syst. Dyn. Rev. **14**(1), 3–29 (1998)

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

Engelfriet, J., Treur, J.: A temporal model theory for default logic. In: Clarke, M., Kruse, R., Moral, S. (eds.) Proceedings of 2nd European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, ECSQARU'93, pp. 91–96, Springer (1994)

Engelfriet, J., Treur, J.: Temporal theories of reasoning. J. Appl. Non-Class. Logics **5**(1), 97–119 (1995). See also in: MacNish, C., Pearce, D., Pereira L.M. (eds.) Logics in Artifical Intelligence, Proceedings of the 4th European Workshop on Logics in Artificial Intelligence, JELIA'94, pp. 279–299. Springer (1994)

Evans, J.: The heuristic-analytic theory of reasoning: extension and evaluation. Psychon Bull. Rev. **13**(3), 378–395 (2006)

Flavell, J.H.: Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. Am. Psychol. **34**(10), 906–911 (1979)

Furlough, C.S., Gillan, D.J.: Mental models: structural differences and the role of experience. J. Cogn. Eng. Decis. Making **12**(4), 269–287 (2018). https://doi.org/10.1177/1555343418773236
[Crossref]

Galton, A.: Operators versus arguments: the ins and outs of reification. Synthese **150**, 415–441 (2006)

Gavrila, I.S., Treur, J.: A formal model for the dynamics of compositional reasoning systems. In: Cohn, A.G. (ed.) Proceedings of the 11th European Conference on Artificial Intelligence, ECAI'94, pp. 307–311. Wiley, Chichester (1994)

Gentner, D.: Structure-mapping: a theoretical framework for analogy. Cogn. Sci. **7**(2), 155–170 (1983)

Gentner, D., Stevens, A. (eds.) Mental Models. Lawrence Erlbaum Associates, Hillsdale, NJ (1983)

Gibbons, J., Gray, M.: An integrated and experience-based approach to social work education: the Newcastle model. Soc. Work. Educ. **21**(5), 529–549 (2002)

Goldman, A.I.: Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading. Oxford University Press, New York (2006)

Granqvist, P., Kirkpatrick, L.A.: Attachment and religious representations and behavior. In: Cassidy, J., Shaver, P.R. (eds.) Handbook of Attachment: Theory, Research, and Clinical Applications, 2nd edn., pp. 906–933. Guilford, New York (2008)

Greca, I.M., Moreira, M.A.: Mental models, conceptual models, and modelling. Int. J. Sci. Educ. **22**(1), 1–11 (2000)

Halford, G.S.: Children's Understanding: The Development of Mental Models. Lawrence Erlbaum Inc. (1993)

Halloun, I.: Schematic modelling for meaningful learning of physics. J. Res. Sci. Teach. **33**, 1019–1041 (1996)

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'21. Lecture Notes in Computer Science, vol. 12744, pp. 462–475. Springer (2021).

Hesslow, G.: Conscious thought as simulation of behaviour and perception. Trends Cogn. Sci. **6**, 242–247 (2002)

Hesslow, G.: The current status of the simulation theory of cognition. Brain Res. **1428**, 71–79 (2012)

Hofstadter, D.R.: Gödel, Escher, Bach. Basic Books, New York (1979)
[zbMATH]

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

Holyoak, K.J., Monti, M.M.: Relational integration in the human brain: a review and synthesis. J. Cogn. Neurosci. (2020)

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)

Johnson-Laird, P.N.: Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness. Harvard University Press (1983)

Johnson-Laird, P.: The history of mental models. In: Manktelow, K., Chung, M.C. (eds.) Psychology of Reasoning: Theoretical and Historical Perspectives. Psychology Press, New York (2004)

Jonker, C.M., Treur, J.: Analysis of the dynamics of reasoning using multiple representations. In: Gray, W.D., Schunn, C.D. (eds.) Proceedings of the 24th Annual Conference of the Cognitive Science Society, CogSci 2002, pp. 512–517. Lawrence Erlbaum Associates, Inc., Mahwah, NJ (2002)

Jonker, C.M., Treur, J.: Modelling the dynamics of reasoning processes: reasoning by assumption. Cogn. Syst. Res. J. **4**, 119–136 (2003)

Kim, J.: Philosophy of Mind. Westview Press (1996)

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'02, pp. 542–547. Lawrence Erlbaum Associates, Mahwah, NJ (2002)

Koriat, A.: Metacognition and consciousness. In: Zelavo, P.D., Moscovitch, M., Thompson, E. (eds.). Cambridge Handbook of Consciousness. Cambridge University Press, New York (2007)

Kozma, R.B.: Learning with media. Rev. Educ. Res. **61**(2), 179–211 (1991)

Kuang, W.X.Y.: The systematicity and coherence of conceptual metaphor. Foreign Lang. Res. 3 (2003)

Lakoff, G.: The contemporary theory of metaphor. In: Ortony, A. (ed.) Metaphor and Thought, pp. 202–251. Cambridge University Press (1993)

Lakoff, G., Johnson, M.: Metaphors We Live By. University of Chicago Press, Chicago (2003)

Landau, M.J., Meier, B.P., Keefer, L.A.: A metaphor-enriched social cognition. Psychol. Bull. **136**(6), 1045–1067 (2010)

Langan-Fox, J., Code, S., Langfield-Smith, K.: Team mental models. Techniques, methods, and analytic approaches. Hum. Factors **42**(2), 242–271 (2000). https://doi.org/10.1518/001872000779656534

Larbi, E., Mavis, O.: The Use of Manipulatives in mathematics education. J. Educ. Pract. **7**(36), 53–61 (2016)

Leary, D.E. (ed.): Metaphors in the history of psychology. Paperback (ed.) Cambridge University Press, Cambridge (1994)

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

Mayer, R.E.: Models for understanding. Rev. Educ. Res. **59**(1), 43–64 (1989)

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

Meela, P., Yuenyong, C.: The study of grade 7 mental model about properties of gas in science learning through model based inquiry (MBI). In: Proceedings of the International Conference for Science Educators and Teachers, pp. 1–6. AIP Conference Proceedings, vol. 2081(030028). AIP Publishing LLC (2019)

Mercer, J.: Understanding Attachment: Parenting, Child Care, and Emotional Development. Greenwood Publishing Group (2006)

Mohammed, S., Ferzandi, L., Hamilton, K.: Metaphor no more. A 15-year review of the team mental model construct. In: J. Manage. **36**(4), 876–910 (2010). https://doi.org/10.1177/0149206309356804

Meyer, J.-J. Ch, Treur, J. (eds.): Dynamics and Management of Reasoning Processes. Springer (2001)

Nagel, E., Newman, J.: Gödel's Proof. New York University Press, New York (1965)
[zbMATH]

Neilson, D., Campbell, T., Allred, B.: Model-based inquiry: a buoyant force module for high school physics classes. Sci. Teach. **77**(8), 38–43 (2010)

Nini, M.: All on the same page: how Team Mental Models (TMM) increase team performance. CQ Net (2019). https://www.ckju.net/en/dossier/team-mental-models-increase-team-performance

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

Piaget, J.: The Construction of Reality in the Child. Basic Books Inc., New York (1954)

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)

Ponterotto, D.: The cohesive role of cognitive metaphor in discourse and conversation. In: Metaphor and Metonymy at the Crossroads: A Cognitive Perspective, pp. 283–298 (2000)

Refaie, E.E.: Understanding visual metaphor: the example of newspaper cartoons. Vis. Commun. **2**(1), 75–95 (2003)

Rizzolatti, G., Craighero, L.: The mirror-neuron system. Annu. Rev. Neurosci. **27**, 169–192 (2004)

Romero, E., Soria, B.: Cognitive metaphor theory revisited. J. Lit. Semant. **34**(1), 1–20 (2005)

Salter, M.D.: An evaluation of adjustment based on the concept of security. Ph.D. thesis, vol 18, p. 72. University of Toronto Studies, Child Development Series (1940)

Salter Ainsworth, M.D.: Security and attachment. In: Volpe, R. (ed.) The Secure Child: Timeless Lessons in Parenting and Childhood Education, pp. 43–53. Information Age Publishing, Charlotte, NC (2010)

Salter Ainsworth, M.D., Bowlby, J.: Child Care and the Growth of Love. Penguin Books, London (1965)

Schaap-Jonker, H., Corveleyn, J.M.: Mentalizing and religion. Arch. Psychol. Relig. **36**(3), 303–322 (2014)

Schjoedt, U., Stodkilde-Jorgensen, H., Geerts, A.W., Roepstorff, A.: Highly religious participants recruit areas of social cognition in personal prayer. SocCog Affect. Neurosci. **4**, 199–207 (2009)

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

Shannon, S.V.: Using metacognitive strategies and learning styles to create self-directed learners. Inst. Learning Styles J. **1**, 14–28 (2008)

Shih, Y.F., Alessi, S.M.: Mental models and transfer of learning in computer programming. J. Res. Comput. Educ. **26**(2), 154–175 (1993)

Skemp, R.R.: The Psychology of Learning Mathematics. Penguin Books, Harmondsworth (1971)

Smorynski, C.: The incompleteness theorems. In: Barwise, J. (ed.) Handbook of Mathematical Logic, vol. 4, pp. 821–865. North-Holland, Amsterdam (1977)

Sterling, L., Beer, R.: Metainterpreters for expert system construction. J. Log. Program. **6**, 163–178 (1989)

Treur, J.: On the use of reflection principles in modelling complex reasoning. Int. J. Intell. Syst. **6**, 277–294 (1991)

Treur, J.: Temporal semantics of meta-level architectures for dynamic control of reasoning. In: Fribourg, L., Turini, F. (ed.) Logic Program Synthesis and Transformation-Meta-Programming in Logic, Proceedings of the Fourth International Workshop on Meta-Programming in Logic, META'94. Lecture Notes in Computer Science, vol. 883, pp. 353–376. Springer (1994)

Treur, J.: Network-Oriented Modeling for Adaptive Networks: Designing Higher-Order Adaptive Biological, Mental and Social Network Models. Springer Nature (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: Proceedings of the 7th International Congress on Information and Communication Technology, ICICT'21. Advances in Intelligent Systems and Computing, vol. 235, pp. 263–274. Springer Nature (2021b)

Treur, J.: Self-modeling networks using adaptive internal mental models for cognitive analysis and support processes. In: Proceedings of the 9th International Conference on Complex Networks and Their Applications, vol. 2. Studies in Computational Intelligence, vol. 944, pp. 260–274. Springer (2021c)

Treur, J.: Mental models in the brain: on context-dependent neural correlates of mental models. Cogn. Syst. Res. **79**, 83–90 (2021d)

Treur, J.: Modeling the emergence of informational content by adaptive networks for temporal factorisation and criterial causation. Cogn. Syst. Res. **68**, 34–52 (2021e)

Treur, J., Van Ments, L. (eds.): Mental Models and their Dynamics, Adaptation and Control: A Self-Modeling Network Modeling Approach. Springer, Cham, Switzerland (2022) (this volume)

Van Gog, T., Paas, F., Marcus, N., Ayres, P., Sweller, J.: The mirror neuron system and observational learning: implications for the effectiveness of dynamic visualizations. Educ. Psychol. Rev. **21**(1), 21–30 (2009)

Van Ments, L., Treur, J.: A higher-order adaptive network model to simulate development of and recovery from PTSD. In: Proceedings of the 11the International Conference on Computational Science, ICCS'21. Lecture Notes in Computer Science, vol. 12743, pp. 154–166. Springer (2021a)

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., Klein, J., Roelofsma, P.H.M.P.: A second-order adaptive network model for shared mental models in hospital teamwork. In: Nguyen, N.T., et al. (eds.) Proceedings of the 13th International Conference on Computational Collective Intelligence, ICCCI'21. Lecture Notes in AI, vol. 12876, pp. 126–140. Springer Nature (2021)

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. In: Treur and Van Ments, 2022, Chap. 11 (this volume) (2022)

Vosniadou, S., Ortony, A. (eds.): Similarity and Analogical Reasoning. Cambridge University Press, New York (1989)

Weyhrauch, R.W.: Prolegomena to a theory of mechanized formal reasoning. Artif. Intell. **13**, 133–170 (1980) [MathSciNet][zbMATH]

Williams, D.: Predictive minds and small-scale models: Kenneth Craik's contribution to cognitive science. Philos. Explor. **21**(2), 245–263 (2018a)

Williams, D.: The mind as a predictive modelling engine: generative models, structural similarity, and mental representation. Ph.D. thesis, University of Cambridge, UK. (2018)

Williams, L.E., Huang, J.Y., Bargh, J.A.: The scaffolded mind: higher mental processes are grounded in early experience of the physical world. Eur. J. Soc. Psychol. **39**(7), 1257–1267 (2009)

Whitaker, K.J., Vendetti, M.S., Wendelken, C., Bunge, S.A.: Neuroscientific insights into the development of analogical reasoning. Dev. Sci. **21**, e12531 (2018). https://doi.org/10.1111/desc.12531

Yi, M.Y., Davis, F.D.: Developing and validating an observational learning model of computer software training and skill acquisition. Inf. Syst. Res. **14**(2), 146–169 (2003)

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_2

2. Bringing Networks to the Next Level: Self-modeling Networks for Adaptivity and Control of Mental Models

Jan Treur[1] ✉
(1) Social AI Group Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
Networks provide an intuitive, declarative way of modeling with a wide scope of applicability. In many cases also adaptivity of a network plays a role, which easily leads to less declarative and transparent forms of modeling by using algorithmic or procedural descriptions for the adaptation processes. This chapter addresses this by exploiting the notion of self-modeling network that has been developed recently. Using that, adaptivity is obtained by adding a self-model to a given base network, with states that represent part of the base network's structure. This adds a next level to the base network, resulting in a two-level network. This construction can easily be iterated to obtain more levels so that multiple orders of adaptation can be covered as well. This brings networks to a next level in more than one way. In particular, a three-level self-modeling network can be used to integrate dynamics, adaptivity and control in one network. It is shown how this can be used to design network models for mental model handling.

Keywords Network-oriented modeling – Self-modeling network – Network reification – Adaptive network model – Controlled adaptation

2.1 Introduction

Recently, within network science new modeling approaches have been developed that can be used to model networks with not only the dynamics of their nodes or states but also the adaptation of their structure and the control of that adaptation. In particular, in this chapter it will be addressed how both within-network dynamics (dynamics of the node states) for networks and adaptivity of the network structure can be addressed using self-modeling networks (Treur 2020a, b).

Self-modeling networks are networks that include a self-model for part of their own network structure in the form of nodes that represent certain network structure characteristics such as connection weights or excitability thresholds. Any (base) network can be extended by including such a self-model, which can be considered to be at a next level, compared to the base network; this self-modeling step is also called network reification. This construction for networks in particular relates to a long-standing tradition in other areas of AI, namely that of meta-programming and metalevel architectures; e.g., Bowen and Kowalski (1982) Demers and Malenfant (1995), Sterling and Shapiro (1996), Sterling and Beer (1989), Weyhrauch (1980). Having such self-models within a network enables to model adaptation of the network structure by the within-network dynamics of the self-model representing this network structure. As the latter can be specified by declarative means in the form of mathematical relations and functions, also adaptivity of the network structure can be specified in a similar declarative manner. To support the modeler, a dedicated software environment is available that also applies to self-modeling networks; see (Treur 2020b, Chap. 9).

In this chapter, the perspective pointed out above will be illustrated in more detail. First in Sect. 2.2 the network-oriented modeling approach based on self-modeling networks will be briefly introduced. In Sect. 2.3 it is discussed how various adaptation principles from the neuroscientific literature can be modeled. Next, in Sect. 2.4 it will be illustrated by an example of a second-order adaptive network model for emotion regulation dysfunction. In Sect. 2.5 it is shown how a self-modeling network model can be used to obtain a computational network model for mental model handling according to the three-level cognitive architecture described in Van Ments and Treur (2021); see also Van Ments and Treur (2022). Finally, Sect. 2.6 is a discussion.

2.2 Modeling Adaptivity by Self-modeling Networks

In this section, the network-oriented modeling approach by self-modeling networks used is briefly introduced in two steps.

2.2.1 Network-Oriented Modeling

As in this approach *nodes Y* in a network have activation values $Y(t)$ that are dynamic over time t, they serve as state variables and will usually be simply called *states*. For these dynamics, the states are considered to affect each other by the connections within the network. Following Treur (2016, 2020b), a basic *network structure* is characterised by:

- **Connectivity characteristics**
  Connections from a state *X* to a state *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 $\boldsymbol{\omega}_{X_i,Y} X_i(t)$ on *Y* from its incoming connections from states $X_1, \ldots, X_k$
- **Timing characteristics**
  Each state *Y* has a *speed factor* $\boldsymbol{\eta}_Y$ defining how fast it changes upon given impact.

Here, the states X_i and *Y* have activation levels $X_i(t)$ and $Y(t)$ that vary (often within the [0, 1] interval) over time, described by real numbers t. The dynamics of such networks are described by the following difference (or differential) equations that incorporate in a canonical manner the network characteristics $\boldsymbol{\omega}_{X,Y}$, $\mathbf{c}_Y(..)$, $\boldsymbol{\eta}_Y$:

$$Y(t+\Delta t) = Y(t) + \boldsymbol{\eta}_Y \left[\mathbf{c}_Y(\boldsymbol{\omega}_{X_1,Y} X_1(t), \ldots, \boldsymbol{\omega}_{X_k,Y} X_k(t)) - Y(t)\right] \Delta t \quad (2.1)$$

for any state *Y* and where $X_1, \ldots, X_k$ are the states from which *Y* gets its incoming connections. The Eq. (2.1) are useful for simulation purposes and also for analysis of properties of the emerging behaviour of such network models. The overall combination function $\mathbf{c}_Y(..)$ for state *Y* is taken as the weighted average of some of the available basic combination functions $\mathbf{c}_j(..)$ by specified weights $\boldsymbol{\gamma}_{j,Y}$, and parameters $\boldsymbol{\pi}_{1,j,Y}$, $\boldsymbol{\pi}_{1,j,Y}$ of $\mathbf{c}_j(..)$, for *Y*:

$$\mathbf{c}_Y(V_1, \cdots, V_k) = \frac{\boldsymbol{\gamma}_{1,Y}\mathbf{c}_1(V_1, \ldots, V_k) + \ldots + \boldsymbol{\gamma}_{m,Y}\mathbf{c}_m(V_1, \ldots, V_k)}{\boldsymbol{\gamma}_{1,Y} + \ldots + \boldsymbol{\gamma}_{m,Y}} \quad (2.2)$$

Such Eqs. (2.1), (2.2) are hidden in the dedicated software environment that can be used for simulation and analysis; see Treur (2020b, Chap. 9). This software environment is freely downloadable from URL.

https://www.researchgate.net/project/Network-Oriented-Modeling-Software.

Combination functions are similar to the functions used in a static manner in the deterministic Structural Causal Model perspective described, for example, in Pearl (2000), Wright (1921), Mooi et al. (2013). However, in the Network-

Oriented Modelling approach described here they are used in a dynamic manner. For example, Pearl (2000, p. 203), denotes nodes by V_i and combination functions by f_i (although he uses a different term for these functions). In the following quote he points at the issue of underspecification concerning aggregation of multiple connections, as in the often used graph representations the specification of combination functions f_i for nodes V_i, is lacking:

> Every causal model M can be associated with a directed graph (…) This graph merely identifies the endogenous and background variables that have a direct influence on each V_i; it does not specify the functional form of f_i. (Pearl 2000, p. 203)

Therefore, in addition to graph representations for connectivity, at least aggregation in terms of combination functions has to be addressed, as indeed is done for the way network models are considered here, in order to avoid this problem of underspecification. That is the reason why aggregation in terms of combination functions is part of the definition of the network structure, in addition to connectivity in terms of connections and their weights and timing in terms of speed factors.

As part of the software environment, a large number > 5 of useful basic combination functions are included in a Combination Function Library, and also a facility to easily indicate any function composition of any available basic combination functions in the library. For a few examples of basic combination functions, see Table 2.1. Here V_1, …,V_k are variables for the single impacts.

Table 2.1 Examples of basic combination functions from the library

Name	Notation	Formula	Parameters
Identity	$\mathbf{id}(V)$	V	–
Scaled sum	$\mathbf{ssum}_{\boldsymbol{\lambda}}(V_1, \ldots, V_k)$	$\frac{V_1+\cdots+V_k}{\lambda}$	Scaling factor $\boldsymbol{\lambda}$
Euclidean	$\mathbf{eucl}_{\boldsymbol{n},\boldsymbol{\lambda}}(V_1, \ldots, V_k)$	$\sqrt[n]{\frac{V_1^n+\cdots+V_k^n}{\lambda}}$	Order $\boldsymbol{n}$ Scaling factor $\boldsymbol{\lambda}$
Advanced logistic	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, \ldots, V_k)$	$\left[\frac{1}{1+e^{-\sigma(V_1+\cdots+V_k-\tau)}} - \frac{1}{1+e^{\sigma\tau}}\right](1+e^{-\sigma\tau})$	Steepness $\boldsymbol{\sigma} > 0$ Threshold $\boldsymbol{\tau}$
Stepmod	$\mathbf{stepmod}_{\boldsymbol{\rho},\boldsymbol{\delta}}(V_1, \ldots, V_k)$	0 if t mod $\boldsymbol{\rho} < \boldsymbol{\delta}$, else 1	Repetition $\boldsymbol{\rho}$ Duration $\boldsymbol{\delta}$

The above concepts (the characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{j,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$) enable to design network models and their dynamics in a declarative manner, based on mathematically defined functions and relations for them. Note that for each state

Y, all characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{j,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$ mentioned above affect the activation level of Y, as also can be seen from Eqs. (2.1) and (2.2). Each of these characteristics provide that influence in their own way from a specific role, either for connectivity, for aggregation or for timing. Below, this observation will also turn out useful in the context of self-models to address adaptivity.

2.2.2 Using Self-modeling Networks to Model Adaptive Networks

Realistic network models are usually adaptive: often some of their network characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{j,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$ change over time. For example, for mental networks often the connections are assumed to change by Hebbian learning (Hebb 1949) and for social networks, it is often assumed that connections between persons change, for example through a bonding by homophily principle (McPherson et al. 2001; Pearson et al. 2006; Sharpanskykh and Treur 2014).

Adaptive networks are often modeled in a hybrid manner by considering two different types of separate models that interact with each other: a network model for the base network and its within-network dynamics, and a numerical model for the adaptivity of (some of) the network structure characteristics of the base network. The latter dynamic model is usually specified in a format outside the context of network modeling: in the form of some adaptation-specific procedural or algorithmic programming specification used to run the difference or differential equations underlying the network adaptation process. In contrast, by including *self-models*, a network-oriented conceptualisation similar to what was described above, can also be applied to obtain adaptive networks as well by using a declarative description based on mathematically defined functions and relations for them.

The approach using self-models was inspired in a metaphorical sense by the more general idea of self-referencing or 'Mise en abyme', sometimes also called 'the Droste-effect' after the famous Dutch chocolat brand who uses this effect in packaging and advertising of their products since 1904. For some examples, see Fig. 2.1. For more explanation, see for example, https://en.wikipedia.org/wiki/Mise_en_abyme, https://en.wikipedia.org/wiki/Droste_effect. This effect occurs in art when within artwork a small copy of the same artwork is included. This can be applied graphically in paintings or photographs, or in sculptures. Also, it is sometimes used within literature (story-within-the-story), theater (theater-within-the-theater), or movies (movie-within-the-movie).

Fig. 2.1 Three examples of the Mise en abyme or Droste-effect
http://michel.parpere.pagesperso-orange.fr/pedago/voc/mise%20en%20abyme.htm
https://www.instagram.com/culturfemale/.
https://www.instagram.com/p/CCYmVLMpGPo/

This idea is applied to model adaptation for a network by adding *self-models* to it as introduced in Treur (2020a, b). This leads to *self-modeling networks*, also called *reified* networks. This works through the addition of new states to the network (called *self-model states*) which represent network characteristics by network states. Then the impacts of these characteristics on a state Y as mentioned above can be modelled as impacts from such self-model states. This brings the impacts from these characteristics on a state Y in the standard form of a network model where via connections nodes affect other nodes.

More specifically, adding a self-model for a base network is done in the way that for some of the states Y of the base network and some of the network structure characteristics for connectivity, aggregation and timing (i.e., some from $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{j,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$), additional network states $\mathbf{W}_{X,Y}$, $\mathbf{C}_{j,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ (*self-model states* or *reification states*) are introduced and connected to other states:

(a) **Connectivity self-model**

- Self-model states $\mathbf{W}_{X,Y}$ are added representing connectivity characteristics, in particular connection weights $\boldsymbol{\omega}_{X,Y}$

(b) **Aggregation self-model**

- Self-model states $\mathbf{C}_{j,Y}$ are added representing aggregation characteristics, in particular combination function weights $\boldsymbol{\gamma}_{j,Y}$

- Self-model states $\mathbf{P}_{i,j,Y}$ are added representing aggregation characteristics, in particular combination function parameters $\boldsymbol{\pi}_{i,j,Y}$

(c) **Timing self-model**

- Self-model states $\mathbf{H}_Y$ are added representing timing characteristics, in particular speed factors $\boldsymbol{\eta}_Y$.

Note that the names using the letters **W**, **C**, **P** and **H** can also be chosen in a different manner. For example, for combination function parameter self-model states often names are used that refer to the specific parameter, for example, **T** for excitability threshold parameter $\boldsymbol{\tau}$, and **M** for persistence parameter $\boldsymbol{\mu}$. The step of adding a self-model to a base network is also called *network reification* and the resulting self-modeling network is sometimes called a *reified network*. If such self-model states are dynamic, they describe adaptive network characteristics. In a graphical 3D-format, such self-model states are depicted at a next level (also called *self-model level* or *reification level*), where the original network is at a *base level*. As an example, the weight $\boldsymbol{\omega}_{X,Y}$ of a connection from state X to state Y can be represented (at a next level) by a self-model state named $\mathbf{W}_{X,Y}$ (e.g., for an objective representation) or $\mathbf{RW}_{X,Y}$ (e.g., for a subjective representation).

Having self-model states to model an adaptation principle in a network-oriented manner is only a first step. To fully model a certain adaptation principle by a self-modeling network, the dynamics of each self-model state itself and its effect on a corresponding target state Y have to be specified in a network-oriented manner by the three general standard types of network characteristics (a) *connectivity*, (b) *aggregation*, and (c) *timing*:

(a) **Connectivity for the self-model states in a self-modeling network**

For the self-model states, their *connectivity* in terms of their incoming and outgoing connections has two different functions:

- **Effectuating its special effect from its specific role**
  The *outgoing downward connections* from the self-model states $\mathbf{W}_{X,Y}$, $\mathbf{C}_{j,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ to state Y represent the specific impact (their special effect from their specific role) each of these self-model states has on Y. These downward impacts are standard per role, and make that the adaptive values $\mathbf{W}_{X,Y}(t)$, $\mathbf{C}_{j,Y}(t)$, $\mathbf{P}_{i,j,Y}(t)$, $\mathbf{H}_Y(t)$ at t are actually used for the adaptive characteristics of the base network in Eqs. (2.1) and (2.2).
- **Indicating the input for the adaptation principle as specified in (b)**

The *incoming upward or leveled connections* to a self-model state are used to specify the *input* needed for the particular adaptation principle that is addressed.

(b)

Aggregation for the self-model states in a self-modeling network

For the self-model states, their aggregation characteristics have one main aim:

- **Expressing the adaptation principle by a mathematical function**

  For the *aggregation* of the incoming impacts for a self-model state, provided as indicated in a), a specific combination function is chosen *to express the adaptation principle* in a declarative mathematical manner.

(c)

Timing for the self-model states in a self-modeling network

For the self-model states, their timing characteristics have one main aim:

- **Expressing the adaptation speed for the adaptation principle by a number**

  Finally, like any other state, self-model states have their own *timing* in terms of speed factors. These speed factors are used as the means to express the adaptation speed.

As a base network extended by including a self-model is also a network model itself, as has been illustrated in Treur (2020b, Chap. 10), this self-modeling construction can easily be applied iteratively to include self-models of multiple self-modeling (or reification) levels. This can provide higher-order adaptive network models, and has turned out quite useful to model, for example, within Cognitive Neuroscience plasticity and metaplasticity (e.g., Abraham and Bear 1996; Garcia 2002; Magerl et al. 2018; Robinson et al. 2016) in a unified form by a second-order adaptive mental network with three levels, one base level and a first- and a second-order self-model level for plasticity and metaplasticity, respectively, as shown in Treur (2020b, Chap. 4).

In the current chapter, the notion of a multi-level self-modeling network model will be illustrated by two examples in Sects. 2.4 and 2.5 from which the latter addresses a higher-order adaptive network model for mental model handling that illustrates how the generic cognitive architecture for mental model handling discussed in Van Ments and Treur (2021, 2022) can be formalized in a computational manner using a self-modeling network.

2.3 Modeling Adaptation Principles

In this section, it will be shown how the modeling approach for self-modeling network models described in Sect. 2.2 can be applied to model adaptation principles as found in empirical sciences. When self-model states are changing over time in a proper manner, this offers a useful method to model any adaptation principle. This does not only apply to first-order adaptive networks, but also to second- or higher-order adaptive networks, for example to model control by using second-order self-models.

2.3.1 First-Order Self-models for First-Order Adaptation Principles

Within Cognitive Neuroscience literature, two types of (first-order) adaptation are often considered, one for connection weights and one for intrinsic neuronal properties; for example, as described in Chandra and Barkai (2018):

> Learning-related cellular changes can be divided into two general groups: modifications that occur at synapses and modifications in the intrinsic properties of the neurons. While it is commonly agreed that changes in strength of connections between neurons in the relevant networks underlie memory storage, ample evidence suggests that modifications in intrinsic neuronal properties may also account for learning related behavioral changes. (Chandra and Barkai 2018, p. 30).

In this chapter for these two types of adaptivity, two examples of first-order adaptation principles are considered: Hebbian Learning for connection weights and Excitability Modulation for the excitability threshold of states.

2.3.1.1 The Hebbian Learning Adaptation Principle

A well-known adaptation principle of the first type (addressing adaptive connectivity) is Hebbian Learning (Hebb 1949), which can be explained by:

> When an axon of cell A is near enough to excite B and repeatedly or persistently (2.3)
> takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.' (Hebb 1949, p. 62)

This is sometimes simplified (neglecting the phrase 'as one of the cells firing B') to:

What fires together, wires together. (Shatz 1992; Keysers and Gazzola 2014) (2.4)

Within a self-modeling network, this can be modeled by using a *connectivity self-model* based on self-model states $\mathbf{W}_{X,Y}$ representing connection weights $\boldsymbol{\omega}_{X,Y}$. These self-model states need incoming and outgoing connections to let them function within the network. To incorporate the 'firing together' part, for the self-model's connectivity, incoming connections from the connected states X and Y to $\mathbf{W}_{X,Y}$ are used; see Fig. 2.2 (upward arrows in blue). These upward connections have weight 1 here. Also a connection from $\mathbf{W}_{X,Y}$ to itself with weight 1 is used to model persistence of the learnt effect; in pictures they are usually left out. In addition, an outgoing connection from $\mathbf{W}_{X,Y}$ to state Y is used to indicate where this self-model state $\mathbf{W}_{X,Y}$ has its effect; see in Fig. 2.2 the (pink) downward arrow. The downward connection indicates that at the base level the value of $\mathbf{W}_{X,Y}$ is actually used for the connection weight of the connection from X to Y.

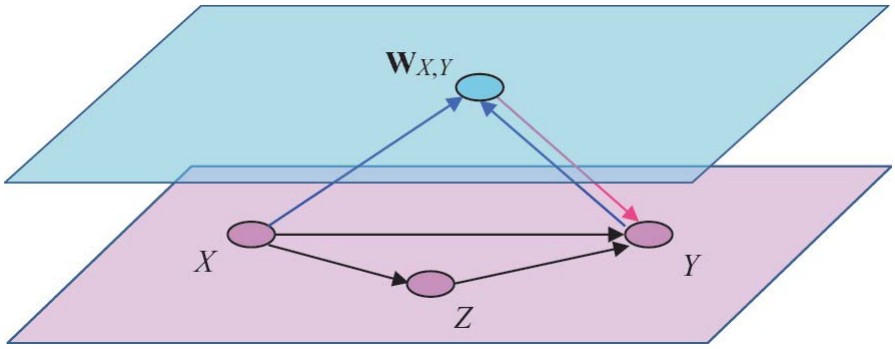


Fig. 2.2 Connectivity characteristics of the self-model for the Hebbian Learning adaptation principle

For the *aggregation characteristics* of the self-model, one of the options for a learning rule is defined by combination function $\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W)$ from Table 2.2. This first-order adaptation principle will be illustrated both by the example adaptive network model discussed in Sect. 2.4 and by the example network for mental model handling in Sect. 2.5.

Table 2.2 Combination functions for self-models modeling first- and second-order adaptation principles used here in the illustrative examples

Name and self-model state	Combination functions	Variables and parameters
Hebbian learning $\mathbf{W}_{X,Y}$	$\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W) = V_1V_2(1-W) + \boldsymbol{\mu}W$	V_1,V_2 activation levels of the connected states W activation level of self-model state $\mathbf{W}_{X,Y}$ $\boldsymbol{\mu}$ persistence factor

Name and self-model state	Combination functions	Variables and parameters
Excitability modulation $\mathbf{T}_Y$	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, ..., V_k)$	$V_1, ..., V_k$ single impacts from base states
Exposure accelerates Adaptation $\mathbf{H}_{\mathbf{WT}_Y}$	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, ..., V_k)$	$V_1, ..., V_k$ single impacts from base states and first-order self-model states

2.3.1.2 The Excitability Modulation Adaptation Principle

Although connectivity adaptation is most often addressed in the literature, it more recently has been pointed out that also other characteristics can be made adaptive, such as excitability thresholds. For example, the following quote indicates that synaptic activity induces long-lasting modifications in excitability of neurons:

> Long-lasting modifications in intrinsic excitability are manifested in changes (2.5)
> in the neuron's response to a given extrinsic current (generated by synaptic activity or applied via the recording electrode). (Chandra and Barkai 2018, p. 30)

For more literature on this form of learning or adaptation (called here the Excitability Modulation adaptation principle), see, for example, Aizenman and Linden (2000), Daoudal and Debanne (2003), Debanne et al. (2019), Lisman et al. (2018), Titley et al. (2017), Zhang and Linden (2003). As here the adaptation depends on activation of a base state Y and the base states (here X, Z) from which it gets its incoming connections, this can be modeled in a self-modeling network in a similar form as above, but this time using a self-model state $\mathbf{T}_Y$; see Fig. 2.3.

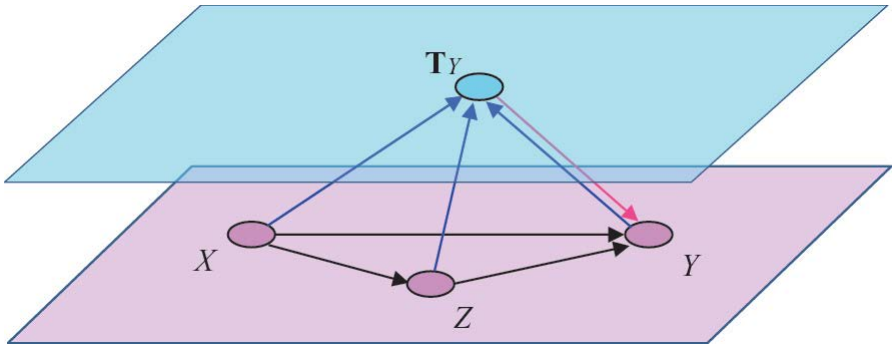


Fig. 2.3 Connectivity characteristics of a self-model for the Excitability Modulation adaptation principle

In this case, based on literature as referred above it is assumed that exposure enhances excitability, which means that it decreases the excitability threshold.

To achieve this, for the self-model state $\mathbf{T}_Y$ a monotonically increasing combination function can be used, while the connection weights from X, Y, Z to $\mathbf{T}_Y$ are negative; examples of monotonically increasing combination functions are the logistic sum functions and the Euclidean function (with odd order $\boldsymbol{n}$) from Table 2.1. In this case, the (pink) downward connection from $\mathbf{T}_Y$ to Y indicates that the value of $\mathbf{T}_Y$ is used for the threshold value of the logistic sum function of base state Y. This first-order adaptation principle will be illustrated by the example network for mental model handling in Sect. 2.5.

2.3.2 Second-Order Self-models for Second-Order Adaptation Principles

The two first-order adaptation principles discussed in Sect. 2.3.1 refer to what in neuroscientific literature is called *plasticity*. It was shown how they can be described by a first-order self-model for connectivity or aggregation characteristics of the base network, in this case in particular for the connection weights or the excitability thresholds used in aggregation. For an organism, in some circumstances it is better to learn (and change) fast, but in other circumstances, it is better to stay stable and let what has been learnt in the past persist: the Plasticity Versus Stability Conundrum (Sjöström et al. 2008, p. 773). Under which circumstances and to which extent such plasticity actually takes place is controlled by a form of socalled *metaplasticity*; e.g., Abraham and Bear (1996), Garcia (2002), Magerl et al. (2018), Robinson et al. (2016), Sjöström et al. (2008). Such control can address 'The Plasticity Versus Stability Conundrum' by only making plasticity happen in circumstances when it is important for the person to change and otherwise stabilise it. In literature as mentioned, various studies show how adaptation (as described, for example, by Hebbian learning), is modulated by accelerating the adaptation process or decelerating or even blocking it. Among the reported factors affecting plasticity in such a way are stimulus exposure, activation, previous experiences, and stress. Here we consider in particular three specific second-order adaptation principles for such control of first-order adaptation: the Adaptation Accelerates with Increasing Exposure, Exposure Modulates Persistence, and Stress Reduces Adaptation adaptation principles.

2.3.2.1 The Adaptation Accelerates with Increasing Exposure Adaptation Principle

For example, in (Robinson et al. 2016) the following compact quote is found summarizing that increasing stimulus exposure makes that the adaptation speed increases:

Adaptation accelerates with increasing stimulus exposure' (2.6)
(Robinson et al. 2016, p. 2).

This indeed describes a form of metaplasticity that controls the speed of adaptation (learning rate). This principle can be modeled by a (dynamic) second-order self-model for timing characteristics (speed factors) of a first-order self-model for the first-order adaptation. Such a second-order is based on self-model states $\mathbf{H}_{\mathbf{W}_{X,Y}}$ or $\mathbf{H}_{\mathbf{T}_Y}$ for adaptive learning speed of any of the two types of (synaptic or intrinsic) learning discussed in Sect. 2.3.1, or $\mathbf{H}_{\mathbf{WT}_Y}$ for both types combined. The principle formulated by (6) indicates that the activation level of these second-order self-model states should depend in a monotonically increasing manner on the activation levels of the base states involved: these base states are *Y* itself and the states *X*, *Z* from which *Y* gets an incoming connection. This makes that the connectivity of this timing self-model (for both forms of learning) is as shown in Fig. 2.4: the (positive, blue) upward connections from the base states *X, Y* and *Z* to the self-model state $\mathbf{H}_{\mathbf{WT}_Y}$ are used to express the part of the principle in (6) referring to 'stimulus exposure'. For the aggregation, for $\mathbf{H}_{\mathbf{WT}_Y}$, a Euclidean combination function (with odd order ***n***) or a logistic sum combination function can be used to get the monotonic effect as needed. The (blue) upward connections from $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ (with negative and positive weight, respectively) to the self-model state $\mathbf{H}_{\mathbf{WT}_{X,Y}}$ indicate a counterbalancing effect that makes that the learning speed is limited depending on a high learnt level as represented by a high value of $\mathbf{W}_{X,Y}$ and a low value of $\mathbf{T}_Y$. The downward (pink) connections from $\mathbf{H}_{\mathbf{WT}_Y}$ to $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ indicate that the value of $\mathbf{H}_{\mathbf{WT}_Y}$ is actually used as speed factor for $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$.

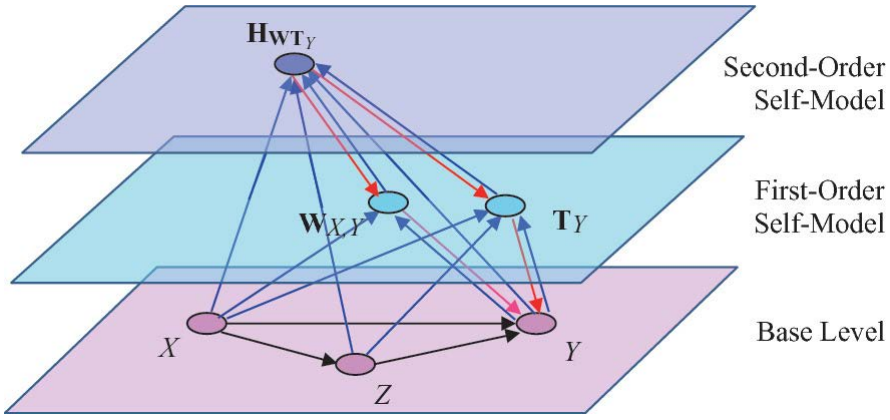


Fig. 2.4 Connectivity of a second-order self-model for the second-order Exposure Accelerates Adaptation adaptation principle for control of first-order self-models for Hebbian Learning and Excitability Modulation

This second-order adaptation principle will be illustrated by the example network for mental model handling discussed in Sect. 2.5. As a small preview for

this, Fig. 2.4 shows how a specific self-modeling network model for mental model handling can be obtained according to the more general three-level cognitive architecture for handling mental models put forward in Van Ments and Treur (2021) with the following three levels:

- **Base level**
  Applying a mental model: the lower (pink) plane.
- **First-order adaptation level**
  Adapting the mental model: the middle (blue) plane.
- **Second-order adaptation level**
  Controlling the adaptation of the mental model: the upper (purple) plane.

2.3.2.2 The Exposure Modulates Persistence Adaptation Principle

A similar perspective can be applied to obtain a principle for modulation of persistence.

Stimulus exposure modulates persistence of adaptation (2.7)

Depending on further context factors, this can be applied in different ways. Reduced persistence can be used in order to get rid of earlier learnt connections that do not apply anymore. However, enhanced persistence can be used to keep what has been learnt. This also is a form of metaplasticity, which can be described by a second-order adaptive network that is modeled using a dynamic second-order *aggregation self-model*, for persistence characteristics of a first-order self-model for the first-order adaptation, based on self-model states $\mathbf{M}_{\mathbf{W}_{X,Y}}$ for an adaptive persistence factor. This second-order adaptation principle will be illustrated by the example adaptive network model discussed in Sect. 2.4 (Fig. 2.5).

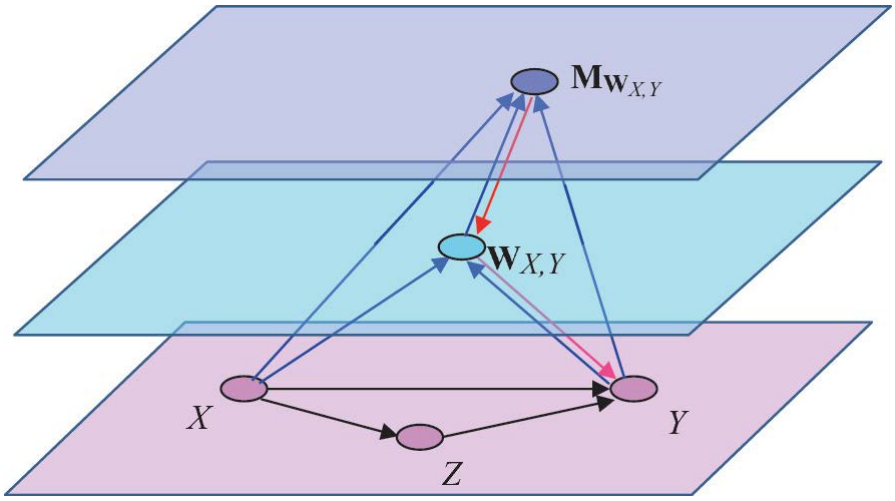


Fig. 2.5 Connectivity of a second-order self-model for the Exposure Modulates Persistence adaptation principle with a first-order self-model for Hebbian learning

2.3.2.3 The Stress Reduces Adaptation Adaptation Principle

In (Garcia 2002) the focus is on the role of stress in reducing or blocking plasticity. Many mental and physical disorders are stress-related, and are hard to overcome due to poor or even blocked plasticity that comes with the stress. Garcia (2002) describes the negative role of stress-related metaplasticity for this, which often becomes a situation that a patient is locked in his or her disorder by that negative pattern. However, he also shows that by some form of therapy this negative cycle may be broken:

> At the cellular level, evidence has emerged indicating neuronal atrophy and cell loss in response to stress and in depression. At the molecular level, it has been suggested that these cellular deficiencies, mostly detected in the hippocampus, result from a decrease in the expression of brain-derived neurotrophic factor (BDNF) associated with elevation of glucocorticoids. (Garcia 2002, p. 629).
>
> …modifications in the threshold for synaptic plasticity that enhances cognitive function is referred here to as 'positive' metaplasticity. In contrast, changes in the threshold for synaptic plasticity that yield impairment of cognitive functions, for example (..) in response to stress (..), is referred to as 'negative' metaplasticity. (Garcia 2002, pp. 630–631).
>
> In summary, depressive-like behavior in animals and human depression are associated with high plasma levels of glucocorticoids that produce 'negative' metaplasticity in limbic structures (…). This stress-related metaplasticity impairs performance on certain hippocampal-dependent tasks. Antidepressant treatments act by increasing expression of BDNF in the hippocampus. This antidepressant effect can trigger, in turn, the suppression of stress-related metaplasticity in hippocampal-hypothalamic pathways thus restoring physiological levels of glucocorticoids.' (Garcia 2002, p. 634).

For this second-order adaptation principle, a picture similar to what is shown in Fig. 2.4 can be drawn, but then for the case that one of the base states represents the stress level and the upward connection of that base state to the **H**-state at the second-order self-model level has a negative weight. This second-order adaptation principle will be illustrated in more detail by the example adaptive network model discussed in Sect. 2.4.

2.4 A Second-Order Adaptive Mental Self-modeling Network Model for Emotion Regulation Dysfunction

In this section, an example of a second-order adaptive network model is described for a mental health context. In such a second-order adaptive network, the base network has its own internal dynamics, but it also uses first-order adaptation principles. Moreover, these first-order adaptation principles themselves change based on second-order adaptation principles.

2.4.1 Design of the Adaptive Network Model for Emotion Regulation Dysfunction

Based on the literature discussed in Sect. 2.3, a second-order adaptive self-modeling network model for plasticity and metaplasticity has been designed with connectivity as shown in Fig. 2.6. Table 2.3 displays the explanations of the states. Here, blocked plasticity for emotion regulation due to stressful feelings is modeled, which leads to dysfunctioning emotion regulation (Garcia 2002).

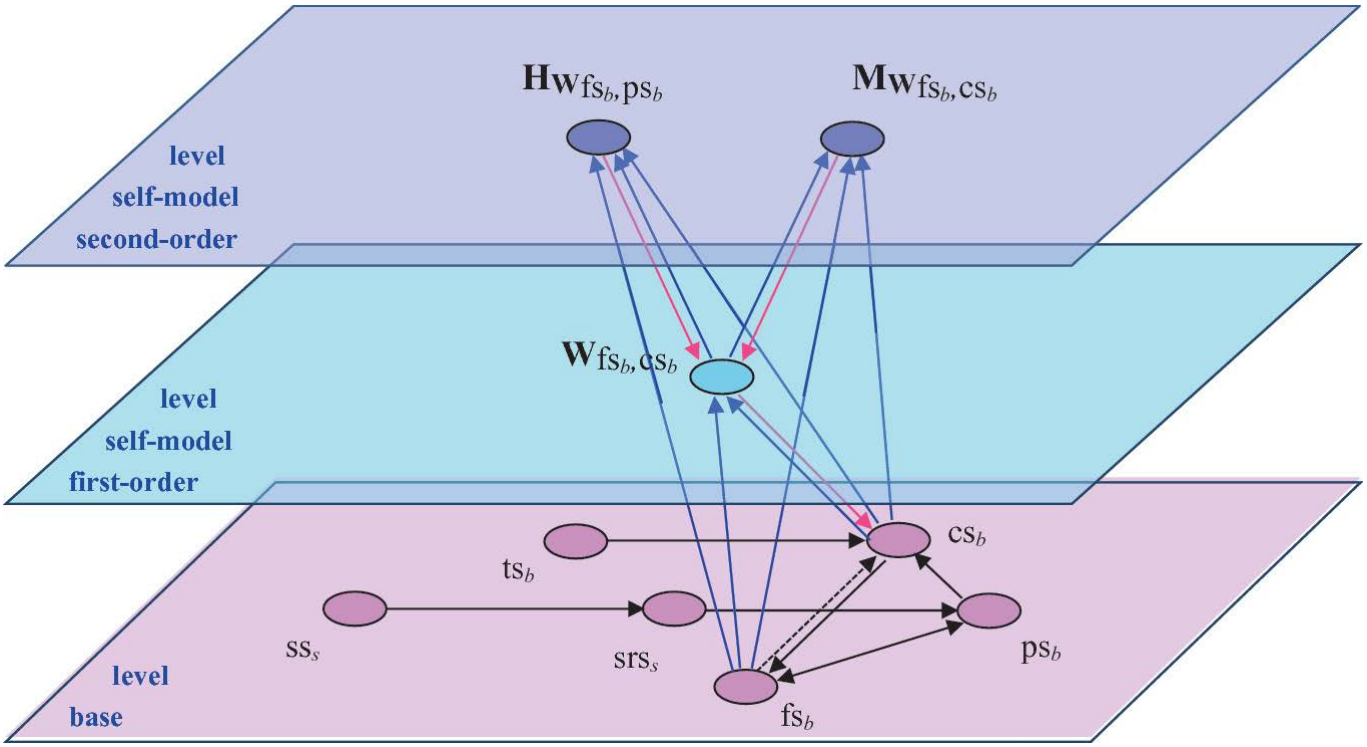


Fig. 2.6 Connectivity of the second-order adaptive network model for plasticity and metaplasticity of emotion regulation with base level (lower plane, pink), first-order self-model level (middle plane, blue) and second-order self-model level (upper plane, purple), and upward connections (blue) and downward connections (red) defining interlevel relations. The dashed arrow indicates the base level connection that is adaptive

Table 2.3 States of the second-order adaptive network model

State nr	State name	Explanation	Level
X_1	ss_s	Sensor state for stimulus s	Base level
X_2	srs_s	Sensory representation state for stimulus s	

State nr	State name	Explanation	Level
X_3	ps_b	Preparation state for emotional response b	
X_4	fs_b	Feeling state for b	
X_5	cs_b	Control state for b	
X_6	ts_b	Therapy state for b	
X_7	$\mathbf{W}_{fs_b,cs_b}$	Representation state for connection weight ω_{srs_s,ps_a}	First-order self-model level
X_8	$\mathbf{H}_{W_{fs_b,cs_b}}$	Representation state for speed factor $\eta_{W_{fs_b,cs_b}}$ for representation state $\mathbf{W}_{fs_b,cs_b}$	Second-order self-model level
X_9	$\mathbf{M}_{W_{fs_b,cs_b}}$	Representation state for persistence factor parameter $\eta_{W_{fs_b,cs_b}}$ for representation state $\mathbf{W}_{fs_b,cs_b}$	

In the base network, s is a stressful stimulus leading to stressful emotional response ps_b and stress feeling fs_b. State cs_b performs stress regulation by its negative outgoing connection, as soon as it is activated. However, to get cs_b activated, the connections from fs_b and ps_b to cs_b play an important role, and such connections have to be learnt and maintained, which may not happen if that plasticity is out of order. Many disorders, both physical and mental, originate from such poor functioning stress regulation. In the example, for the sake of simplicity, we focus on the connection from fs_b to cs_b. The learning process for this connection is modeled by the Hebbian learning principle (see Sect. 2.3.1.1 and Fig. 2.2) represented by first-order self-model state $\mathbf{W}_{fs_b,cs_b}$. This state models a Hebbian learning adaptation principle defined as a combination function by the function $\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W)$ as indicated in Table 2.2.

However, as discussed in Sect. 2.3.2, whether or not learning takes place depends on metaplasticity, modelled here by second-order self-model states $\mathbf{H}_{W_{fs_b,cs_b}}$ and $\mathbf{M}_{W_{fs_b,cs_b}}$ for learning speed and persistence, respectively (see also Sects. 2.3.2.1 and 2.3.2.2 and Figs. 2.4 and 2.5). Note that the **M**-state is an aggregation self-model state for a combination function parameter: the persistence parameter $\boldsymbol{\mu}$ for the combination function $\mathbf{hebb}_{\boldsymbol{\mu}}(..)$ for first-order adaptation state $\mathbf{W}_{fs_b,cs_b}$ for Hebbian learning of the connection from fs_b to cs_b. In principle, the learning will start to work or accelerate when the external stimulus s is sensed through sensor state ss_s. As discussed in Sect. 2.3.2.3, stress creates a negative metaplasticity effect, which corresponds to low values for these second-order self-model states. For example, a value close to 0 for learning speed representation $\mathbf{H}_{W_{fs_b,cs_b}}$ practically blocks the learning, and a value around

0.5 for persistence representation $\mathbf{M}_{\mathbf{W}_{\mathrm{fs}_b,\mathrm{cs}_b}}$ makes that every time unit, around 50% of the learnt effect is lost, which is a dramatic effect if no additional learning takes place.

For most of the states in the designed network model, for the aggregation characteristics $\mathbf{alogistic}_{\sigma,\tau}(..)$ is used as combination function; see Table 2.1. The only exceptions are $\mathbf{W}_{\mathrm{fs}_b,\mathrm{cs}_b}$ which uses $\mathbf{hebb}_{\mu}(..)$ defined by (4) and Table 2.2, and therapy state ts_b which is an external input modeled by $\mathbf{stepmod}_{\rho,\delta}(..)$ which defines an (independent) activation after *duration* $\boldsymbol{\delta}$ and *repetition* of the cycle after a time period $\boldsymbol{\rho}$.

2.4.2 Specification of the Adaptive Network Model for Emotion Regulation Disfunction

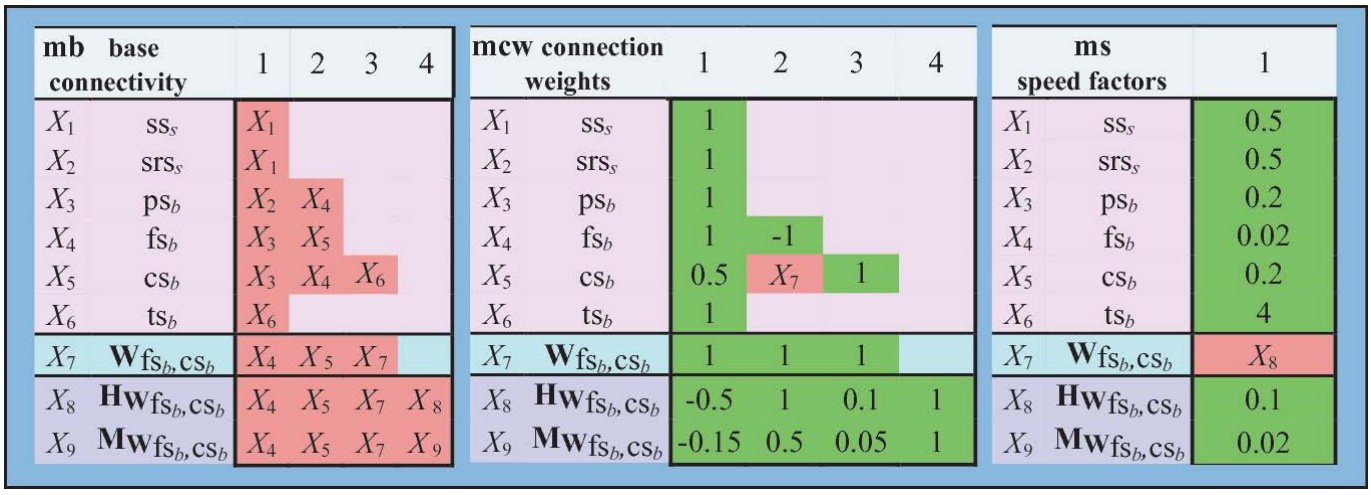
To get a well-defined standard format to specify a design of a self-modeling network, the socalled role matrix format has been introduced in Treur (2020b). In the first place, a specification in that format can be used as a compact but detailed, neat and standardised form of documentation for human use, for example as a basis for communication among designers. It does not only cover all information on connectivity as represented in graphical format in pictures such as shown in Fig. 2.6, but also on all other network characteristics defining a network model, such as $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{j,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$ as discussed in Sect. 2.2. Therefore, a self-modeling network model is fully defined by this specification in role matrix format. For an overview of the five role matrices in relation to the network characteristics, see Table 2.4. In the second place, the standardized format of the role matrices makes it relatively easy to implement a software environment that can use them as input and based on that runs simulations. So, role matrices do not only form a good basis for documentation and communication among humans, they are also a good basis for communication with computers. Such a software environment and how to use it is described in Treur (2020b, Chap. 9); see also Treur (2022a).

Table 2.4 Overview of the role matrices for the different types of network characteristics

	Network characteristics	**Role matrix**	**Notation**
Connectivity characteristics	Base connectivity	**mb**	Picture (upward and horizontal arrows)
	Connection weights	**mcw**	$\boldsymbol{\omega}_{X,Y}$
Aggregation characteristics	Combination function weights	**mcfw**	$\boldsymbol{\gamma}_{j,Y}$
	Combination function parameters	**mcfp**	$\boldsymbol{\pi}_{i,j,Y}$

	Network characteristics	Role matrix	Notation
Timing characteristics	Speed factors	**ms**	$\boldsymbol{\eta}_Y$

In Figs. 2.7 and 2.8 all network characteristics for the designed adaptive network model for emotion regulation dysfunction are specified in the form of role matrices. Role matrix **mb** in Fig. 2.7 specifies the *base connectivity* characteristics. On each row, for the given state (in the left column) it indicates from which states at the same or a lower level it gets an incoming connection (the black and blue arrows in Fig. 2.6). Note that for some of the states a connection from the state itself occurs. The latter applies to all (first- and second-order) self-model states, as can be seen in **mb**. Such connections are usually not depicted in graphical representations such as the one in Fig. 2.6.

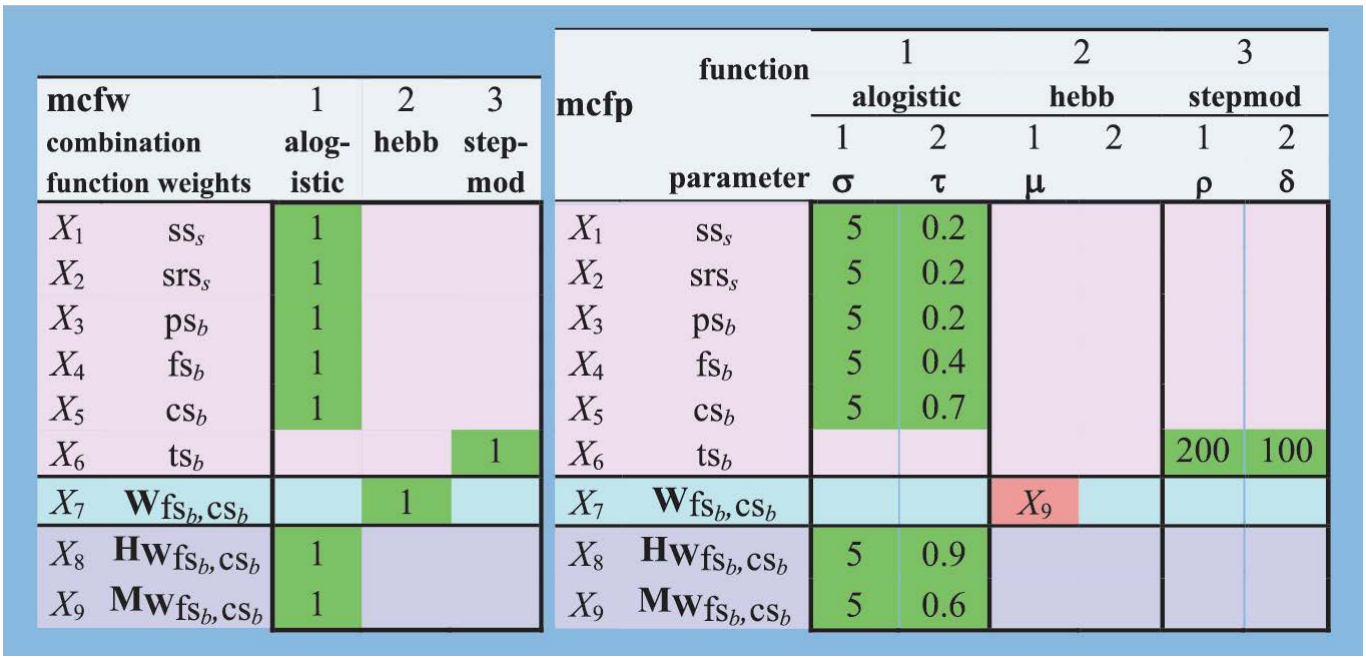


mb base connectivity		1	2	3	4
X_1	ss_s	X_1			
X_2	srs_s	X_1			
X_3	ps_b	X_2	X_4		
X_4	fs_b	X_3	X_5		
X_5	cs_b	X_3	X_4	X_6	
X_6	ts_b	X_6			
X_7	$\mathbf{W}fs_b,cs_b$	X_4	X_5	X_7	
X_8	$\mathbf{Hw}fs_b,cs_b$	X_4	X_5	X_7	X_8
X_9	$\mathbf{Mw}fs_b,cs_b$	X_4	X_5	X_7	X_9

mcw connection weights		1	2	3	4
X_1	ss_s	1			
X_2	srs_s	1			
X_3	ps_b	1			
X_4	fs_b	1	-1		
X_5	cs_b	0.5	X_7	1	
X_6	ts_b	1			
X_7	$\mathbf{W}fs_b,cs_b$	1	1	1	
X_8	$\mathbf{Hw}fs_b,cs_b$	-0.5	1	0.1	1
X_9	$\mathbf{Mw}fs_b,cs_b$	-0.15	0.5	0.05	1

ms speed factors		1
X_1	ss_s	0.5
X_2	srs_s	0.5
X_3	ps_b	0.2
X_4	fs_b	0.02
X_5	cs_b	0.2
X_6	ts_b	4
X_7	$\mathbf{W}fs_b,cs_b$	X_8
X_8	$\mathbf{Hw}fs_b,cs_b$	0.1
X_9	$\mathbf{Mw}fs_b,cs_b$	0.02

Fig. 2.7 Specification of the *connectivity characteristics* and *timing characteristics* of the second-order adaptive network model for emotion regulation dysfunction by role matrices **mb**, **mcw** and **ms**



mcfw combination function weights		1 alog-istic	2 hebb	3 step-mod
X_1	ss_s	1		
X_2	srs_s	1		
X_3	ps_b	1		
X_4	fs_b	1		
X_5	cs_b	1		
X_6	ts_b			1
X_7	$\mathbf{W}_{fs_b,cs_b}$		1	
X_8	$\mathbf{Hw}_{fs_b,cs_b}$	1		
X_9	$\mathbf{Mw}_{fs_b,cs_b}$	1		

mcfp function		1 alogistic		2 hebb		3 stepmod	
		1	2	1	2	1	2
parameter		σ	τ	μ		ρ	δ
X_1	ss_s	5	0.2				
X_2	srs_s	5	0.2				
X_3	ps_b	5	0.2				
X_4	fs_b	5	0.4				
X_5	cs_b	5	0.7				
X_6	ts_b					200	100
X_7	$\mathbf{W}_{fs_b,cs_b}$			X_9			
X_8	$\mathbf{Hw}_{fs_b,cs_b}$	5	0.9				
X_9	$\mathbf{Mw}_{fs_b,cs_b}$	5	0.6				

Fig. 2.8 Specification of the *aggregation characteristics* of the second-order adaptive network model for emotion regulation disfunction by role matrices **mcfw** and **mcwfp**

As an example, in the second row, it is indicated that state X_2 (= srs_s) only has one incoming base connection, from state X_1 (= ss_s). As another example, the seventh row indicates that state X_7 (= $\mathbf{W}_{fs_b,cs_b}$) has incoming base connections from X_4 (= fs_b), X_5 (= cs_b), X_7 (= $\mathbf{W}_{fs_b,cs_b}$) itself, and in that order. This order is important as the Hebbian combination function $\mathbf{hebb}_{\mu}(..)$ used is not symmetric in its arguments. Note that the more informative state names such as ss_s, and so on, in each of the role matrices depicted in Figs. 2.7 and 2.8 are actually not part of the specification as used in the computer, but are only for human understanding. In a similar way the other types of role matrices are defined; see Figs. 2.7 and 2.8: role matrices **mcw** for connection weights, **mcfw** for combination function weights, **mcfp** for combination function parameters, and **ms** for speed factor roles. Here the combination functions selected from the library are specified by **mcf** = [....], for the current example it is **mcf** = [2 3 35]; here the numbers 2, 3, 35 refer to the numbers in the combination function library, where $\mathbf{alogistic}_{,}(\ldots)$ has number 2 and $\mathbf{hebb}\,(\ldots)$ number 3; number 35 is the **stepmod** function used to create independent events. By specifying **mcf** = [2 3 35] for this specific network model they become combination function numbers 1 to 3 as also shown in role matrices **mcfw** and **mcfp**.

Within each role matrix, for an adaptive network characteristic, entries in red cells indicate a reference to the name of another state that as self-model state represents that characteristic, while entries in green cells indicate fixed values for nonadaptive characteristics. In this way the red cells represent the pink

downward connections from the self-model states in pictures as shown in Fig. 2.6, with their specific roles **W**, **H**, **C**, **P** indicated by the type of role matrix: the type of role matrix in which they are represented, defines the roles of the self-model states. For example, in **mcw** the X_7 in the peach-red cell in the row for X_5 defines that the connection weight for the connection to X_5 (= cs_b) from X_4 (= fs_b) is adaptive with value represented by X_7 (= $\mathbf{W}_{fs_b,cs_b}$). So, by specifying X_7 in role matrix **mcw** in that cell, X_7 gets the role of connection weight self-model state for the connection from fs_b to cs_b. Similarly, role matrix **ms** indicates (in peach-red) that X_8 plays the role of the (adaptive) speed factor of X_7, and (in green) that the speed factors of all other states have fixed values.

In Fig. 2.8 the role matrices **mcfw** and **mcfp** are shown for aggregation characteristics in terms of combination function weights and parameters, respectively. Matrix **mcfp** is a 3D matrix with first dimension for the states, second dimension for the (two) combination function parameters and third dimension for the combination functions. For example, in Fig. 2.8 the name X_9 in the red cell in role matrix **mcfp** indicates that the value of the persistence parameter **μ** for X_7 (= $\mathbf{W}_{fs_b,cs_b}$) is adaptive and is represented by the value of state X_9 (= $\mathbb{M}_{W_{fs_b,cs_b}}$). In contrast, the 5 in the first green cell of **mcfp** for X_5 indicates the static value of the steepness of the logistic function for X_5 (= cs_b).

For this example network model, the selection of combination functions from the library for the network is specified by **mcf** = [2 3 35], being **alogistic$_{\sigma,\tau}$(..)**, **hebb$_{\mu}$(..)**, **stepmod$_{\rho,\delta}$(..)**, respectively. So, in terms of (2) for this network it holds c_1(..) = **alogistic$_{\sigma,\tau}$(..)**, c_2(..) = **hebb$_{\mu}$(..)**, c_3(..) = **stepmod$_{\rho,\delta}$(..)**.

2.4.3 Simulations for the Adaptive Network for Emotion Regulation Dysfunction

A number of simulation experiments have been performed using the dedicated software environment for self-modeling network models described in Treur (2020), Chap. 9); see also Treur (2022a). In particular, a scenario is shown here in which the focus was on the effect of the stress level of fs_b on plasticity, following Garcia (2002). In Fig. 2.9 simulation results are shown for the characteristics in Fig. 2.1 and 2.2. Here a person is considered who for some time (before time 0) has had a stressful life, so that the initial values already reflect a high stress level. Initial values were as shown in Table 2.5.

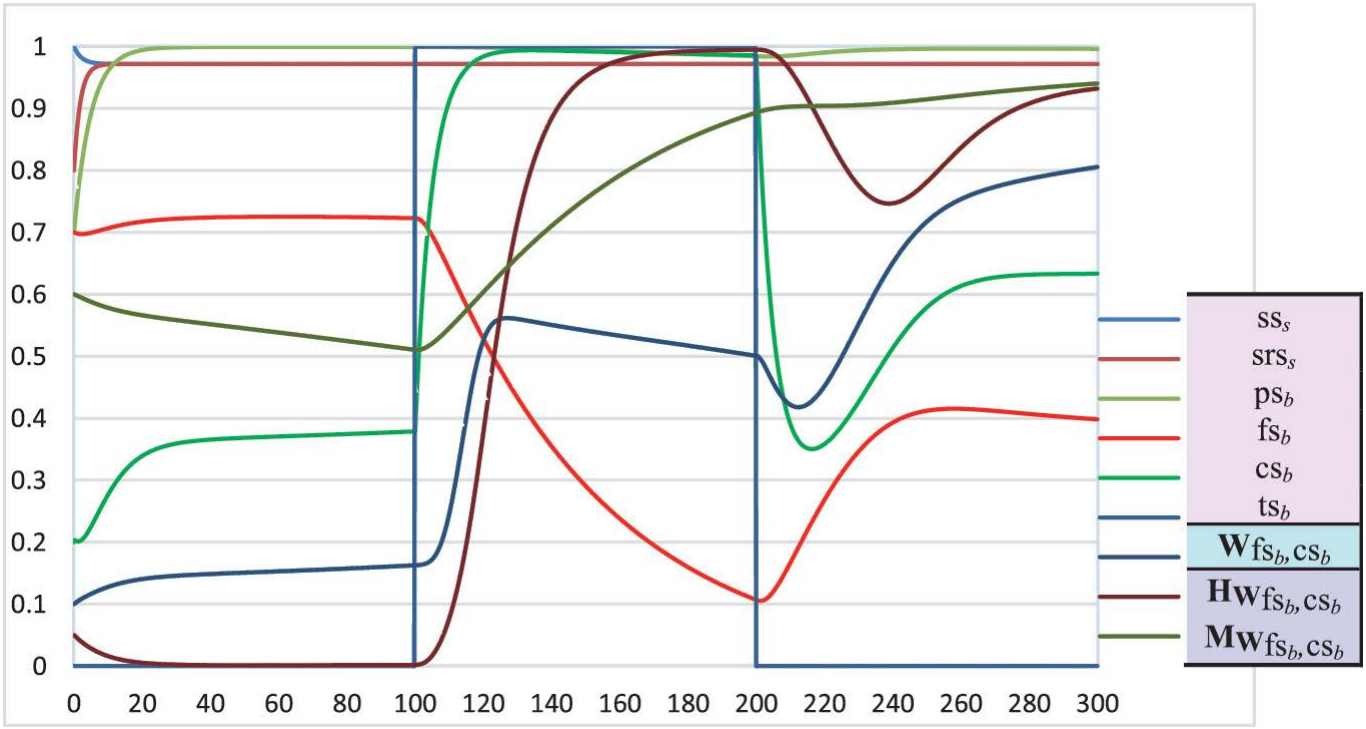


Fig. 2.9 Simulations for the second-order adaptive example network model. First phase (time 0 to 100): a high stress level while metaplasticity blocks plasticity, due to which emotion regulation does not function properly. Second phase (time 100–200): a therapy artifically boosts emotion regulation, which makes metaplasticity unblock the plasticity which in turn makes emotion regulation stronger. Third phase (time 200–300): due to the unblocked plasticity, the emotion regulation is now able to learn further and keep the stress level low without therapy

Table 2.5 Initial values for the simulation

ss_s	srs_s	ps_b	fs_b	cs_b	ts_b	W_{fs_b,cs_b}	$M_{W_{fs_b,cs_b}}$	$M_{W_{fs_b,cs_b}}$
1	0.8	0.7	0.7	0.2	0.1	0.1	0.05	0.6

The graph in Fig. 2.9 shows the activation levels of all states, including how the weight of the connection from fs_b to cs_b represented by W_{fs_b,cs_b} is learnt, and what speed and persistence values are applied for that, represented by $H_{W_{fs_b,cs_b}}$ and $M_{W_{fs_b,cs_b}}$. As can be seen, in the first phase until time 100 a high stress level represented by fs_b (the red line) leads to maintaining low values of $H_{W_{fs_b,cs_b}}$ (the brown line starting at 0.05) and $M_{W_{fs_b,cs_b}}$ (the grey line starting at 0.6); therefore learning is practically blocked in this phase (the blue line for W_{fs_b,cs_b} starting at 0.1 stays low). In the next phase, from time 100 to time 200 a therapy is applied, represented by ts_b, that gives a boost to the activation level of cs_b, which in turn reduces the stress level represented by fs_b. This also increases the speed and persistence of the Hebbian learning (second-order states $H_{W_{fs_b,cs_b}}$ and $M_{W_{fs_b,cs_b}}$ increase to 1 and 0.9, respectively) and because of this, now indeed learning takes place. In the last phase from time 200 to time 300 the therapy has finished,

but as due to the therapy the stress regulation mechanism has been unlocked, now it is able to keep the stress levels low without external help.

2.5 An Example Network Model for Mental Model Handling

In this section it is shown how an adaptive network model to handle a mental model can be designed. This is done based on the three-level cognitive architecture described in Van Ments and Treur (2021, 2022). For this, first an example scenario is described in Sect. 2.5.1. After that in Sects. 2.5.2 and 2.5.3 the network design and its specifation are presented.

2.5.1 An Example Scenario for Mental Model Handling

The scenario used for this section concerns learning of a mental model by a new person in a company who has to learn to recognize a colleague. It goes as follows.

Example Scenario

A new person in a company has to learn to recognize a colleague from only seeing his face; this face is stimulus *s*. Two colleagues a_1 and a_2 are assumed that are options to choose from. Picking one of them is indicated by activation of sensory representation state srs_{a_i}. A belief bs_1 suggests that it is colleague a_1, and a belief bs_2 that it is colleague a_2. These beliefs are only meant indicative (e.g., based on the location at which the person is encountered), but not sufficient to decide for one of them. As the beliefs and *s* are triggered by independent circumstantial factors, for the network model they just happen. Two types of network characteristics are addressed as adaptive: the weights of the connections from sensory representation srs_s to srs_{a_i} and to srs_{a_i}, and the excitability thresholds for sensory representation states srs_{a_i} and srs_{a_i}. The small network consisting of these three base level states together with the first-order self-model states representing the characteristics for connection weights and excitability thresholds form a mental model of our subject for the colleague considered here; the connection defines how strongly (in the mental model in the mind of our subject) the face relates to the name of the person. During the scenario these characteristics are learnt so that over time a better mental model and decision result. Then in future situations an encounter with *s* (also at unexpected locations, such as in a shop or in another town) leads to correct recognition.

2.5.2 Connectivity and Aggregation for the Adaptive Network Model

In this section a second-order adaptive network model is introduced that addresses the above example scenario, according to the three-level cognitive architecture pointed out in Van Ments and Treur (2021, 2022):

- a base level for the base mental model
- a first-order self-model level for adaptation of the mental model
- a second-order self-model for control of the adaptation.

As discussed in Van Ments and Treur (2021, 2022), the general idea is that a mental model is some structure defined by relations. Using the network-oriented modeling approach adopted here, they can be described by a base network where the relations are modeled as connections. As these relations can change, for example by learning, a (first-order) self-model of the base model is used with self-model states that represent the relations of the base level. At the self-model level, these self-model states are also connected by some type of relations that define when and how adaptation can take place. These relations are represented by second-order self-model states in a second-order self-model. These second-order self-model states determine control of the adaptation at the first self-model level. To determine that, they have their own relations for the second-order self-model level. Based on the network-oriented modeling approach used here, all such relations at (and between) the different levels are modeled as connections.

For the example described in Sect. 2.5.1, first in Fig. 2.10 the connectivity of the base level is depicted. The base mental model consists of three states srs_s, srs_{a_i} and srs_{a_i} (in the darker sheded area) and relations between them. Here especially the relations from srs_s to srs_{a_i} and from srs_s to srs_{a_i} are considered, as depicted by dashed arrows in Fig. 2.10.

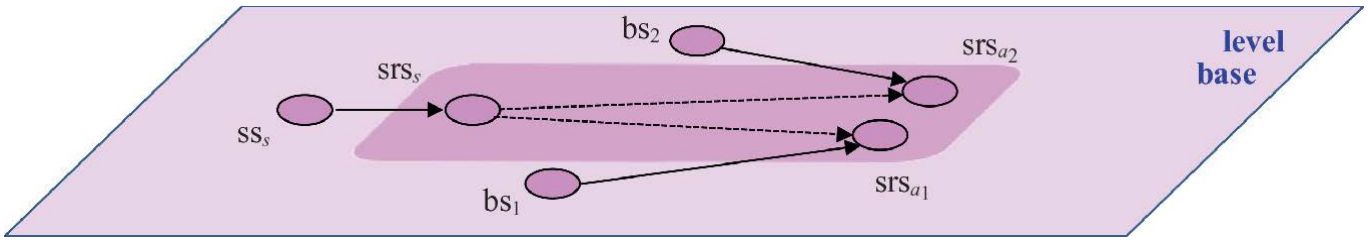


Fig. 2.10 The base level for the example mental model for recognition; the darker shaded part is the mental model, the dashed arrows indicate the connections that are adaptive

After observing the colleague (via ss_s), sensory representation state srs_s gets activated, which, depending on the weights of the connections from srs_s to srs_{a_i} and srs_{a_i} and the excitability thresholds of srs_{a_i} and srs_{a_i} ideally activates one of

srs_{a_i} and srs_{a_i} indicating the correct colleague, without any of the beliefs bs_1 and bs_2 being activated. However, in the beginning it is not that ideal: in the first phase, the activation of the correct belief is needed to be able to make a choice between srs_{a_i} and srs_{a_i}. Due to learning, later on this dependence on the beliefs is not needed anymore as the connection from srs_s to the relevant option is strengthened by this learning and the excitability threshold gets lower for that option.

The following two types of self-model states are used to define the adaptation by learning in the considered mental model:

- **Connectivity self-model states for connection weights**
  The states $\mathbf{W}_{X,Y}$ play the role of connection weight for the adaptive connection from X to Y. They model the Hebbian Learning adaptation principle (Hebb 1949); see also (4) in Sect. 2.3.1.1.
- **Aggregation self-model states for excitability thresholds**
  The states $\mathbf{T}_Y$ play the combination function parameter role for state Y's adaptive excitability threshold $\boldsymbol{\tau}$. They model the Excitability Modulation adaptation principle (Chandra and Barkai 2018); see also (5) in Sect. 2.3.1.2.

Moreover, to control the adaptation, second-order timing self-model states are used:

- **Timing self-model states for connection weight self-model states**
  The second-order self-model states $\mathbf{H}_{\mathbf{WT}_{X,Y}}$ play the role of speed factor for the first-order connection weight self-model states $\mathbf{W}_{X,Y}$ for the adaptive connections from X to Y and for the first-order excitability threshold self-model states $\mathbf{T}_Y$ for the adaptive excitability thresholds of Y. These second-order self-model states model the second-order adaptation principle called Adaptation Accelerates with Increased Exposure (Robinson et al. 2016); see also (6) in Sect. 2.3.2.1.

All in all, this creates a (sub)network for the core mental model and including its adaptation and control based on the following states:

- base state srs_s for the image of the face
- the two base states srs_{a_i} and srs_{a_i} for the options of colleagues and their excitability thresholds
- the two connections (dashed arrows) from srs_s to srs_{a_i} and srs_{a_i} for the options of colleagues with their weights
- the two first-order connectivity self-model states $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$ and $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$ for the weights of these two base connections

- the two first-order aggregation self-model states $\mathbf{T}_{\mathrm{srs}_{a_1}}$ and $\mathbf{T}_{\mathrm{srs}_{a_1}}$ for the excitability thresholds of states srs_{a_i} and srs_{a_i}
- the two second-order timing self-model states $\mathbf{H}_{\mathbf{WT}_{X,Y}}$ for the connectivity self-model states $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$ and $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$ for the weights of these two base connections and for the first-order aggregation self-model states $\mathbf{T}_{\mathrm{srs}_{a_1}}$ and $\mathbf{T}_{\mathrm{srs}_{a_1}}$ for the excitability thresholds of states srs_{a_i} and srs_{a_i}

This core mental model can also be extended by adding the base level belief states to it as well. In a graphical representation of the mental model's *connectivity* in a 3D format, the first-order self-model states are placed in a second (blue) plane, above the (pink) plane for the base mental model, and the second-order self-model states in a third (purple) plane above the second (blue) plane. See Fig. 2.11 and see Table 2.2 for explanations of all states. The following types of connection are used: upward and downward connections, and horizontal leveled connections. Downward connections have a particular effect, as they are effectuating one of the types of adaptive characteristics indicated by their role **W**, **C**, **P** or **H**; see also Sect. 2.2 (Fig. 2.11).

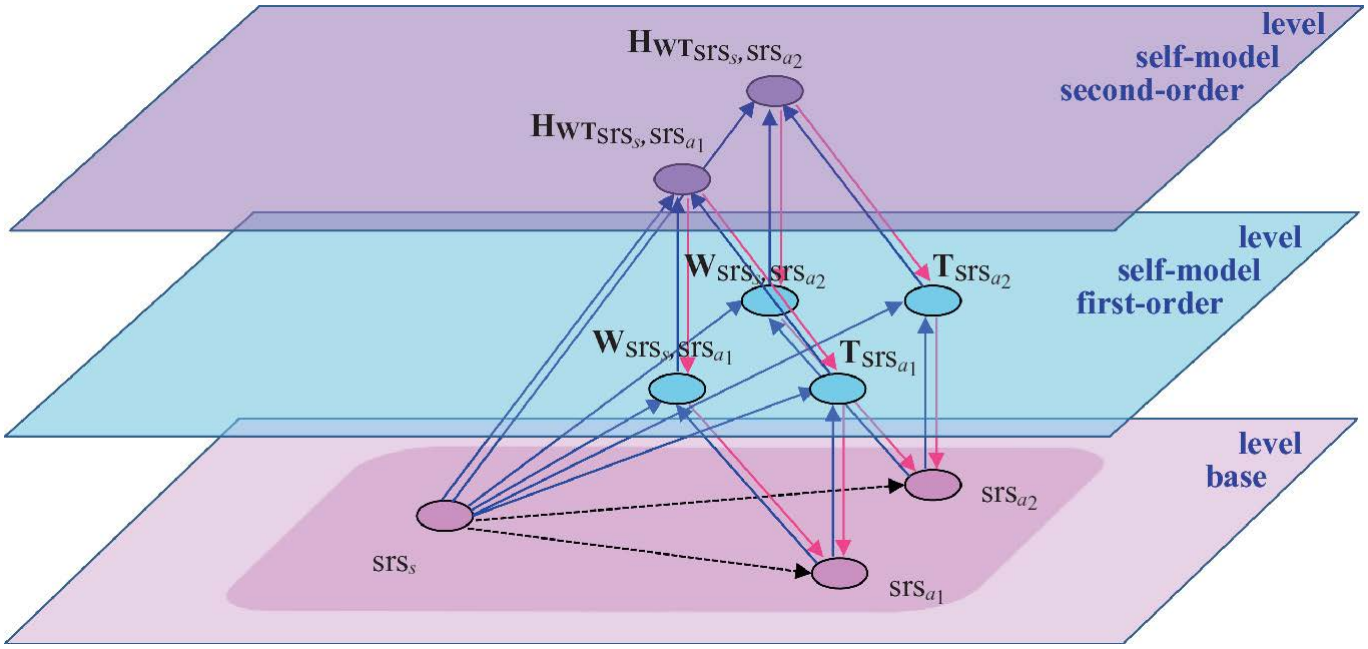


Fig. 2.11 3D representation of the connectivity of the core mental model for recognition, including: (1) *base level* for the face recognition (depicted by the lower, pink plane), (2) *first-order self-model level* (depicted by the middle, blue plane) for the two **W**-states for the weights of the base connections from srs_S to srs_{a_i} and srs_{a_i} and for the two **T**-states for the excitability thresholds for the two base states srs_{a_i} and srs_{a_i}, (3) *second-order self-model level* (depicted by the upper, purple plane) for the two $\mathbf{H}_{\mathbf{WT}}$-states for the speed factors for the **W**-states for the weights of the base connections from srs_S to srs_{a_i} and srs_{a_i} and for the two **T**-states for the excitability thresholds for the two base states srs_{a_i} and srs_{a_i}

For *aggregation*, in the example mental model, for the three base level states the logistic function $\mathbf{alogistic}_{\sigma,\tau}(\ldots)$ is used, and also for the aggregation (excitability threshold) self-model states $\mathbf{T}_{\mathrm{srs}_{a_1}}$ and $\mathbf{T}_{\mathrm{srs}_{a_1}}$. To model Hebbian learning, combination function $\mathbf{hebb}_{\mu}(V_1, V_2, W)$ is applied for the two connectivity (connection weight) self-model states $\mathbf{W}_{X,Y}$ of the mental model; see Table 2.6.

Table 2.6 The states used in the example network model for mental model handling

State nr	name	Explanation
X_1	ss_s	Sensor state for stimulus *s* (seeing a face)
X_2	srs_s	Sensory representation state for stimulus *s*
X_3	bs_1	Belief state 1 (belief that it is Person 1)
X_4	bs_2	Belief state 2 (belief that it is Person 2)
X_5	srs_{a1}	Sensory representation state for recognition as Person 1
X_6	srs_{a2}	Sensory representation state for recognition as Person 2
X_7	$\mathrm{w}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$	First-order self-model state for the weight of the connection from srs_s to srs_{a_i}
X_8	$\mathrm{w}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$	First-order self-model state for the weight of the connection from srs_s to srs_{a_i}
X_9	$\mathbf{T}_{\mathrm{srs}_{a_1}}$	First-order self-model state for the excitability threshold of srs_{a_i}
X_{10}	$\mathbf{T}_{\mathrm{srs}_{a_1}}$	First-order self-model state for the excitability threshold of srs_{a_i}
X_{11}	$\mathrm{HWT}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$	Second-order self-model for the speed factor of the first-order self-model state $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$ for the weight of the connection from srs_s to srs_{a_i} and of the first-order self-model state $\mathbf{T}_{\mathrm{srs}_{a_1}}$ for the excitability threshold of srs_{a_i}
X_{12}	$\mathrm{HWT}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$	Second-order self-model for the speed factor of the first-order self-model state $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$ for the weight of the connection from srs_s to srs_{a_i} and of the first-order self-model state $\mathbf{T}_{\mathrm{srs}_{a_1}}$ for the excitability threshold of srs_{a_i}

2.5.3 Specification of the Example Network Model for Mental Model Handling

The network model for mental model handling was specified by role matrices as shown in Fig. 2.12. As discussed in Sect. 2.4.2, they are **mb** (for the base connection role), **mcw** (for the connection weight role), **ms** (for the speed factor role), **mcfw** (for the combination function weight role), and **mcfp** (for the combination function parameter role).

mb	base connectivity	1	2	3
X_1	ss_s	X_1		
X_2	srs_s	X_1		
X_3	bs_1	X_3		
X_4	bs_2	X_4		
X_5	srs_{a1}	X_2	X_3	
X_6	srs_{a2}	X_2	X_4	
X_7	$\mathbf{W}srs_s, srs_{a1}$	X_2	X_5	X_7
X_8	$\mathbf{W}srs_s, srs_{a2}$	X_2	X_6	X_8
X_9	$\mathbf{T}srs_{a1}$	X_2	X_5	X_9
X_{10}	$\mathbf{T}srs_{a2}$	X_2	X_6	X_{10}
X_{11}	$\mathbf{HWT}srs_s, srs_{a1}$	X_1	X_7	X_9
X_{12}	$\mathbf{HWT}srs_s, srs_{a2}$	X_1	X_8	X_{10}

mcw	connection weights	1	2	3
X_1	ss_s	1		
X_2	srs_s	1		
X_3	bs_1	1		
X_4	bs_2	1		
X_5	srs_{a1}	X_7	0.5	
X_6	srs_{a2}	X_8	0.5	
X_7	$\mathbf{W}srs_s, srs_{a1}$	1	1	1
X_8	$\mathbf{W}srs_s, srs_{a2}$	1	1	1
X_9	$\mathbf{T}srs_{a1}$	-0.2	-0.2	1
X_{10}	$\mathbf{T}srs_{a2}$	-0.2	-0.2	1
X_{11}	$\mathbf{HWT}srs_s, srs_{a1}$	1	-1	1
X_{12}	$\mathbf{HWT}srs_s, srs_{a2}$	1	-1	1

mcfw	combination function weights	1 alo-gistic	2 hebb	3 step-mod
X_1	ss_s			1
X_2	srs_s	1		
X_3	bs_1			1
X_4	bs_2			1
X_5	srs_{a1}	1		
X_6	srs_{a2}	1		
X_7	$\mathbf{W}srs_s, srs_{a1}$		1	
X_8	$\mathbf{W}srs_s, srs_{a2}$		1	
X_9	$\mathbf{T}srs_{a1}$	1		
X_{10}	$\mathbf{T}srs_{a2}$	1		
X_{11}	$\mathbf{HWT}srs_s, srs_{a1}$	1		
X_{12}	$\mathbf{HWT}srs_s, srs_{a2}$	1		

mcfp	function	1 alogistic		2 hebb		3 stepmod	
		1	2	1	2	1	2
	parameter	σ	τ	μ		ρ	δ
X_1	ss_s					50	25
X_2	srs_s	5	0.8				
X_3	bs_1					70	60
X_4	bs_2					50	25
X_5	srs_{a1}	5	X_9				
X_6	srs_{a2}	5	X_{10}				
X_7	$\mathbf{W}srs_s, ps_{a1}$			0.95			
X_8	$\mathbf{W}srs_s, ps_{a2}$			0.95			
X_9	$\mathbf{T}srs_{a1}$	5	0.4				
X_{10}	$\mathbf{T}srs_{a2}$	5	0.4				
X_{11}	$\mathbf{HWT}srs_s, srs_{a1}$	5	0.8				
X_{12}	$\mathbf{HWT}srs_s, srs_{a2}$	5	0.8				

ms	speed	1
X_1	ss_s	2
X_2	srs_s	0.5
X_3	bs_1	2
X_4	bs_2	2
X_5	srs_{a1}	0.2
X_6	srs_{a2}	0.5
X_7	$\mathbf{W}srs_s, srs_{a1}$	0.3
X_8	$\mathbf{W}srs_s, srs_{a2}$	0.3
X_9	$\mathbf{T}srs_{a1}$	0.07
X_{10}	$\mathbf{T}srs_{a2}$	0.07
X_{11}	$\mathbf{HWT}srs_s, srs_{a1}$	0.2
X_{12}	$\mathbf{HWT}srs_s, srs_{a2}$	0.2

initial values		
X_1	ss_s	0
X_2	srs_s	0
X_3	bs_1	0
X_4	bs_2	0
X_5	srs_{a1}	0
X_6	srs_{a2}	0
X_7	$\mathbf{W}srs_s, ps_{a1}$	0.3
X_8	$\mathbf{W}srs_s, ps_{a2}$	0.3
X_9	$\mathbf{T}srs_{a1}$	0.8
X_{10}	$\mathbf{T}srs_{a2}$	0.8
X_{11}	$\mathbf{HWT}srs_s, srs_{a1}$	0
X_{12}	$\mathbf{HWT}srs_s, srs_{a2}$	0

Fig. 2.12 Role matrices specification for the example network for mental model handling

For a designed model a list of combination functions used is specified by **mcf** = [….], for the current example it is **mcf** = [2 3 35]; here the numbers 2, 3, 35 refer to the numbers in the combination function library, where $\mathbf{alogistic}_{\sigma,\tau}(\ldots)$ has number 2 and $\boldsymbol{\omega}_{X_i,Y} X_i(t)$ number 3; number 35 is the stepmod function used to create independent events. Figure 2.12 shows all role matrices for the adaptive network model addressing mental model handling.

Role matrix **mb** specifying *base connectivity* indicates at each row for the indicated state X_j from which states it gets incoming connections from the same or a lower level. For example, the 5th row indicates for state X_5 (= srs_{a_i}) two incoming base connections, one from state X_2 (= srs_s), and one from state X_3 (= bs_1). For another example, row 7 indicates that state X_7 (= $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$) has incoming base connections from X_2 (= srs_s), X_5 (= srs_{a_i}) and from X_7 itself in that order; this ordering is crucial since the Hebbian combination function $\mathbf{hebb}(\ldots)$ used for this state X_7 (= $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$) is not symmetric in its three arguments, as can be seen in Table 2.2.

The other four role matrices are as follows: role matrices **mcw** for the connection weight role and **ms** for the speed factor role, and role matrices **mcfw** for the combination function weight role and **mcfp** for the combination function parameter role (see Fig. 2.12). Within each of these non-base role matrices cell entries in peach-red cells show the name of a state (at a higher level) that as self-model state represents an adaptive characteristic; in contrast, entries in green cells indicate static values for nonadaptive characteristics. Therefore, as seen in Fig. 2.12 the peach-red cells in **mcw** and **mcfp** refer to the (self-model) states X_7 to X_{10}. For example, in role matrix **mcw** the indication X_7 in the peach-red cell at row 5 and column 1 specifies that the value of state X_7 represents the weight of the connection from srs_s to srs_{a_i} (as indicated in **mb**). Unlike this, the 1 in green cell at row 7, column 1 of **mcw** shows the nonadaptive value of weight of the connection from X_2 (= srs_s) to X_7 (= $\mathbf{W}_{\mathrm{srs}_s,\mathrm{srs}_{a_1}}$). In role matrix **mcfp** specifying the combination function parameter role, in the peach-red cell at row 5 and column 2 it is specified that the actual value for the excitability threshold of srs_{a_i} is represented by the value of self-model state X_9 (= $\mathbf{T}_{\mathrm{srs}_{a_1}}$). More explanation of this specification format and how it is used to automatically generate simulations can be found in Treur (2020a, b, 2022a).

A simulation scenario and a more extensive analysis from an informational viewpoint of this example network for mental model handling can be found in Sects. 16.6 and 16.7 (in Chap. 16 of this volume) of Treur (2022e).

2.6 Discussion

For many domains network models provide an intuitive, declarative way of modeling supported by graphical representations. Connections between nodes in a network can be used as a format to model different types of relations occurring in real-world situations. Once network models are represented within a computer, these relations can be used for some types of computational processes to generate *within-network dynamics*, such as simulation processes or reasoning processes or some (other) types of analyses. However, as relations in real-world domains often change over time themselves too, network models for realistic situations also need facilities to change their structure. This is called dynamics *of* networks, or *network adaptation*. It has turned out that socalled self-modeling networks enable to model such changes in network structure relatively easily. These are networks that include nodes that represent specific network characteristics of the network itself and in this way form a self-model of part of the network's own structure. By using a self-model, the adaptation of the network structure can be modeled as within-network dynamics of this self-model.

The above applies in particular to mental models as they also are usually described by relations and these relations can change, for example, due to learning. In this chapter it has been illustrated how self-modeling networks can be applied as useful means to design computational models for mental model handling. This does not only concern the use of a mental model and the adaptation of it as indicated above, but also control over the adaptation. In this way a three-level self-modeling network architecture is obtained that fits very well to the global cognitive architecture found in Van Ments and Treur (2021); see also Van Ments and Treur (2022). As for self-modeling networks a dedicated software environment is available to easily simulate them, this can be used to relate this cognitive architecture to a computational network model by which simulation experiments can be performed.

To analyse the scope of applicability of this network-oriented modeling approach based on self-modeling networks, following Ashby (1960) in Treur (2017), Sect. 2.3.1 it has been shown that any (state-determined) dynamical system as defined in Ashby (1960) and also used in Port and van Gelder (1995) can be described by a set of first-order differential equations, and conversely; see also Treur (2021a, d). Moreover, in Treur (2017), Sect. 2.3.2 it has also been shown how any set of first-order differential equations can be (re)modeled by a network model. These methods can also be applied to adaptive processes: any description of an adaptation process by a dynamical system or by first-order differential equations can be rewritten as a self-model in a self-modeling network. This has been used in Treur 2021a) to show that any adaptive

dynamical system can be modeled as a self-modeling network; see also Treur (2022d).

Finally, analysis of stationary points and equilibria for self-modeling network models has been addressed in Treur (2016, Chap 12, 2020b, Chap. 11–14, 2021a, 2022b). Validation and parameter tuning has been addressed (Treur 2016, Chap. 14, 2022c).

References

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

Aizenman, C.D., Linden, D.J.: Rapid, synaptically driven increases in the intrinsic excitability of cerebellar deep nuclear neurons. Nat. Neurosci. **3**, 109–111 (2000)
[Crossref]

Anten, J., Earle, J., Treur, J.: An Adaptive computational network model for strange loops in political evolution in society. In; Proceedings of the 20th International Conference on Computational Science, ICCS'20, vol. 2, pp. 604–617. Lecture Notes in Computer Science, vol. 12138. Springer (2020)

Ashby, W.R.: Design for a Brain, 2nd extended edn. Chapman and Hall, London. First edition, 1952 (1960)

Bowen, K.A., Kowalski, R.: Amalgamating language and meta-language in logic programming. In: Clark, K., Tarnlund, E. (eds.) Logic Programming, pp. 153–172. Academic Press, New York (1982)

Carley, K.M.: Inhibiting adaptation. In Proceedings of the 2002 Command and Control Research and Technology Symposium, pp. 1–10. Naval Postgraduate School, Monterey, CA

Carley, K.M.: Destabilization of covert networks. Comput. Math. Organiz. Theor. **12**, 51–66 (2006)
[Crossref]

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

Daoudal, G., Debanne, D.: Long-term plasticity of intrinsic excitability: learning rules and mechanisms. Learn. Memory **10**, 456–465 (2003)
[Crossref]

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

Demers, F.N., Malenfant, J.: Reflection in logic, functional and objectoriented programming: a Short Comparative Study. In: IJCAI'95 Workshop on Reflection and Meta-Level Architecture and their Application in AI, pp. 29–38 (1995)

Fessler, D.M.T., Clark, J.A., Clint, E.K.: Evolutionary psychology and evolutionary anthropology. In: The Handbook of Evolutionary Psychology, D.M. Buss edn., pp. 1029–1046. Wiley (2015)

Fessler, D.M.T., Eng, S.J., Navarrete, C.D.: Elevated disgust sensitivity in the first trimester of pregnancy: evidence supporting the compensatory prophylaxis hypothesis. Evol. Hum. Behav. **26**(4), 344–351 (2005)

[Crossref]

Garcia, R.: Stress, metaplasticity, and antidepressants. Curr. Mol. Med. **2**, 629–638 (2002)
[Crossref]

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

Hofstadter, D.R.: Gödel, Escher, Bach. Basic Books, New York (1979)
[zbMATH]

Keysers, C., Gazzola, V.: Hebbian learning and predictive mirror neurons for actions, sensations and emotions. Philos. Trans. r. Soc. Lond. B Biol. Sci. **369**, 20130175 (2014)
[Crossref]

Levy, D.A., Nail, P.R.: Contagion: a theoretical and empirical review and reconceptualization. Genet. Soc. Gen. Psychol. Monogr. **119**(2), 233–284 (1993)

Lisman, J., Cooper, K., Sehgal, M., Silva, A.J.: Memory formation depends on both synapse-specific modifications of synaptic strength and cell-specific increases in excitability. Nat. Neurosci. **21**, 309–314 (2018)
[Crossref]

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

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

Mooij, J.M., Janzing, D., Schölkopf, B.: From differential equations to structural causal models: the deterministic case. In: Nicholson, A., Smyth, P. (eds.) Proceedings of the 29th Annual Conference on Uncertainty in Artificial Intelligence (UAI-13), pp. 440--448. AUAI Press (2013)

Pearl, J.: Causality. Cambridge University Press (2000)

Pearson, M., Steglich, C., Snijders, T.: Homophily and assimilation among sport-active adolescent substance users. Connections **27**(1), 47–63 (2006)

Port, R.F., Van Gelder, T.: Mind as Motion: Explorations in the Dynamics of Cognition. MIT Press, Cambridge, MA (1995)

Robinson, B.L., Harper, N.S., McAlpine, D.: Meta-adaptation in the auditory midbrain under cortical influence. Nat. Commun. **7**, 13442 (2016)
[Crossref]

Rojas, R.: Neural Networks. Springer, Berlin (1996)
[Crossref]

Sharpanskykh, A., Treur, J.: Modelling and analysis of social contagion in dynamic networks. Neurocomputing **146**, 140–150 (2014)
[Crossref]

Shatz, C.J.: The developing brain. Sci. Am. **267**, 60–67 (1992). https://doi.org/10.1038/scientificamerican0992-60

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

[Crossref]

Sterling, L., Shapiro, E.: The Art of Prolog, Chap. 17, pp. 319–356. MIT Press (1996)

Sterling, L., Beer, R.: Metainterpreters for expert system construction. J. Log. Program. **6**, 163–178 (1989)
[Crossref]

Titley, H.K., Brunel, N., Hansel, C.: Toward a neurocentric view of learning. Neuron **95**, 19–32 (2017)
[Crossref]

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions. Springer (2016)
[Crossref]

Treur, J.: On the applicability of network-oriented modeling based on temporal-causal networks: why network models do not just model networks. J Inf Telecommun **1**(1), 23–40 (2017)

Treur, J.: Multilevel network reification: representing higher order adaptivity in a network. In: Aiello, L., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L. (eds.), Proceedings of the 7th International Conference on Complex Networks and their Applications, ComplexNetworks'18, vol. 1. Studies in Computational Intelligence, vol. 812, pp. 635–651. Springer (2018)

Treur, J.: Modeling higher-order adaptivity of a network by multilevel network reification. Netw. Sci. **8**, S110–S144 (2020a)
[Crossref]

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

Treur, J.: On the Dynamics and Adaptivity of Mental Processes: Relating Adaptive Dynamical Systems and Self-Modeling Network Models by Mathematical Analysis. Cognitive Systems Research, vol. 70, pp. 93–100 (2021a)

Treur, J.: Equilibrium Analysis of within-network dynamics: from linear to nonlinear aggregation. In: Nguyen, N.T., et al. (eds.) Proceedings of the 13th International Conference on Computational Collective Intelligence, ICCCI'21. Lecture Notes in AI, vol. 12876, pp. 94–110. Springer (2021b)

Treur, J.: With a Little help: a modeling environment for self-modeling network models. In: Treur, J., van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: a Self-Modeling Network Modeling Approach, Chap. 17. Springer, Switzerland (this volume) (2022a)

Treur, J.: Where is this leading me: stationary point and equilibrium analysis of self-modeling network models. In: Treur, J., van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: a Self-Modeling Network Modeling Approach, Chap. 18. Springer, Switzerland (this volume) (2022b)

Treur J.: Does this suit me: validation of self-modeling network models by parameter tuning. In: Treur, J., van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: A Self-Modeling Network Modeling Approach, Chap. 19. Springer, Switzerland (this volume) (2022c)

Treur, J.: How far do self-modeling network models reach: relating them to adaptive dynamical systems. In: Treur, J., van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: A Self-Modeling Network Modeling Approach, Chap 20. Springer, Switzerland (this volume) (2022d)

Treur, J.: How the brain creates emergent information by the development of mental models: an analysis from the perspective of temporal factorisation and criterial causation. In: Treur, J., van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: a Self-Modeling Network Modeling Approach, Chap 16. Springer, Switzerland (this volume) (2022e)

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

Van Ments, L., Treur, J.: Dynamics, adaptation and control for mental models: a cognitive architecture. In: Treur, J., van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: a Self-Modeling Network Modeling Approach, Chap. 1. Springer, Switzerland (this volume) (2022)

Weyhrauch, R.W.: Prolegomena to a theory of mechanized formal reasoning. Artif. Intell. **13**, 133–170 (1980) [MathSciNet][Crossref]

Wright, S.: Correlation and causation. J. Agric. Res. **20**, 557–585 (1921)

Zhang, W., Linden, D.J.: The other side of the engram: experience-driven changes in neuronal intrinsic excitability. Nat. Rev. Neurosci. **4**, 885–900 (2003)
[Crossref]

Part II
Self-Modelling Network Models for Mental Models in Individual Processes

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_3

3. On Becoming a Good Driver: Modeling the Learning of a Mental Model

Raj Bhalwankar[1, 2] ✉ and Jan Treur[2] ✉
(1) Work and Social Psychology Department, Maastricht University, Maastricht, Netherlands
(2) Social AI Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
Mental models play a crucial role in individual and organizational learning. The adaptive mental processes involved in the development of mental models are addressed here by integrating psychological and neurological theories on mental models and the learning involved. An adaptive network model has been designed for these processes and used for simulations addressing a case study of learning to drive a car. The developed model may be valuable for different ends, like in improving individual and organizational learning, in designing virtual pedagogical agents, enhancing driver safety, and self-car-driving systems.

Keywords Mental model – Learning – Instruction – Observational

3.1 Introduction

Understanding the processes of learning and instruction has been a major topic in the field of education. Mental models play a very important role in understanding the construction of knowledge and the actions of individuals interacting with their environment (Kim 2004; Senge 1990). Mental models are indicated to be useful in learning, retrieving, and problem-solving in Olsen (1992), Rouse and Morris (1986). Also, they are essential in understanding complex tasks that may have observable and hidden interactions (Fein et al.

1993). For organizations, it is an important construct that serves to enhance organizational learning as they capture an individual's comprehension of a specific domain in their mind. Many questions relating to mental models are still not fully answered; for example: How do we develop mental models? How can we better facilitate their development? How is their applicability for developing intelligent systems? The study reported here is meant to contribute some further answers to such questions.

Part of the literature on mental models addresses the role of mental models of devices or device models in operating equipment; e.g., Kieras and Bovair (1984). Psychologists believe that such device models aid learning, recall, and transfer of the procedures of operating the devices in complex systems (Kieras and Bovair 1984). While most studies have focused on the theoretical principles about the aid that a mental model provides, very few have empirically tested its benefits when used in the context of learning and operating devices (Fein et al. 1993). Greca and Moreira (2000) acknowledged that the modeling process is indeed very complex and that the scientific understanding of it is still quite poor. They also mentioned our inability to identify the different mental models students have in a given domain and about the specific mental models that they construct. Furthermore, there have been only a few attempts to simulate the development of mental models. For example Devi et al. (1996), utilized AI modeling based on a production rule modeling format to simulate the students' construction of energy models in physics. They mention that they faced difficulties in incorporating the more sophisticated modeling processes in students. There appears to be a lack of research into how these models develop, especially when engaged in learning to operate any device, say a common complex device such as a car.

Network-Oriented Modeling is a useful method to represent complex real-world processes concerning human beings. It has been applied to model networks for a wide range of processes, especially biological, mental and social processes and interactions thereof (Treur 2016, 2020). It can be used to model individual adaptive mental processes as interactions of mental states while their connections change based on principles of network adaptation such as Hebbian learning (Hebb 1949). The adaptive Network-Oriented Modeling approach described in Treur (2020) can be used to model such adaptive processes in a relatively easy manner. The mechanisms describing the development of mental models can be modeled by an adaptive network using (dynamic) states and (adaptive) connections between them. This has been used to model and simulate the development of a mental model, illustrated for a case study on the role of mental models while learning to drive a car.

Such a driver's mental model and how it develops would be beneficial in designing virtual pedagogical agents in the educational technology sector, and in

improving the symbiotic relationship between driver and the increasing adaptive automation in cars thus increasing driver safety. It can be also used to enhance self-driving systems in cars. Moray (1999) argued that quantitative models of the psychological processes can help (1) ergonomists in designing the systems, (2) predict behavior in a particular situation like vehicle automation, and (3) generalize the feedback to other domains.

The chapter starts by providing a literature overview in Sect. 3.2, in which a brief literature overview of the existing research related to the topic is discussed. Then, in Sect. 3.3 the design of the developed adaptive network model with its various parts is discussed. The simulation of the example scenario is discussed in Sect. 3.4. Section 3.5 addresses discussion and conclusions.

3.2 Literature Overview

The history of Mental Models may have started with Kenneth Craik's book 'The Nature of Explanation' (Craik 1943); he writes:

> if the organism carries a 'small-scale model' of external reality and of its possible actions within its head, (…) 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.

Johnson-Laird (1983) took a slightly different logic-inspired approach to Craik's notion by his research on mental models conceptualized as a reasoning mechanism in the working memory. Gentner and Stevens (1983) and Barquero (1995) took an instructional approach to mental models. The main role of a mental model in the instructional approach is considered to allow its builder to explain and make predictions about the physical system represented by it. A popular approach from educational psychology called 'model-based learning' says that learning occurs when people construct meaningful representations, such as coherent mental models that represent and communicate subjective experiences, ideas, thoughts, and feelings of a person (Mayer 1989). According to Buckley (2000):

> Model-based learning is a dynamic, recursive process of learning by building mental models. It incorporates the formation, testing, and subsequent reinforcement, revision, or rejection of mental models of some phenomenon.

Rumelhart et al. (1986) proposed a cognitive architecture consisting of two cognitive processing modules: an interpretation network (activation of schemas

and pattern matching) and a model of the world (creating expectations about states of the world and predicting outcomes). This cognitive architecture was supported by the notion of model-based reinforcement learning from the field of neuroscience. In model-based reinforcement learning, candidate actions are evaluated with regard to the expected outcomes in accordance with a world model (Doll et al. 2012). Moreover, it is important to note that research on reinforcement learning in neuroscience has found neural evidence indicating the existence of mental models. Reinforcement learning provides neural evidence for a cognitive module responsible for the construction of mental models but it does not explain how this construction occurs. It is the functional approach to model-based learning proposed by Buckley (2000) which explains that creation of mental models works in a stepwise enrichment process which begins with activation of preconceptions about the specific domain knowledge which is integrated into the initial working mental model and then it is tested to check whether it fulfills the requirement of the task. In case of insufficiency to its purpose, it is modified by means of accretion, tuning, and restructuring after which it is tested again. This process goes on until a final 'target model' is developed which works well (Seel 1991).

The modeling process has been described in a variety of ways. Halloun (1996) defined it as:

> "... the learning of a series of steps to identify only those salient elements of a system, and to evaluate, according to distinct rules, the chosen model."

Sutton (1996) stated the modeling process as 'the learning of a new language' that would allow for another perception of the description of the phenomena. Nersessian (1995) explained it as:

> "... the integrated reasoning process which uses analogical and visual modeling, as well as thought experiments in the creation and transformation of the informal representations of a problem."

Bandura and Walters (1977) through their famous experiment demonstrated the transmission of aggression through the imitation of aggressive role models. They indicated that from observing another person one forms an idea and that most human behavior is learned observationally through the process of modeling. Several papers published, for example (Gentner and Stevens 1983) support the idea that people have mental models inside their heads and these mental models must be computationally simulatable. In De Kleer and Brown

(1983) it is suggested that these simulations, both the mental and the computationally implemented ones, involve two steps:

1. the envisioning of the system, including a topological representation of the system components, the possible states of each of the components, and the structural relations between these components
2. the running or execution of the causal model based on basic operational rules and on general scientific principles.

Research by Benbassat (2014) on role models in medical education suggests that role modeling is an important part of clinical training when it involves demonstration of skills, feedback and reflective imitation. According to Van Gog et al. (2009), a large body of empirical research has confirmed that learning from 'expert' models, either by observing them solving problems 'live' or watching modeling examples on video or by even studying a written account of their problem-solving process, is a very effective way for acquiring both motor and cognitive skills. These studies also mention that learning from such expert models is more effective for novice learners than learning by solving the equivalent problems. In line with the idea of mirroring (Rizzolatti and Craighero 2004), Hurley (2008) suggests that the processes of motor control, mirroring, and mental simulation rely on shared neural circuits. Yi and Davis (2003) state that model-based training affects the learning of the task or phenomena by influencing one or more of the four observational learning processes:

1. Attention
2. Retention
3. Production
4. Motivation.

For observational learning to take place, the learner is supposed to convert the retained symbolic memory of the observed actions into overt actions to learn the desired response.

Based on this literature, in next section it is discussed how the mental process of the development of mental models can be modeled by an adaptive network model.

3.3 An Adaptive Network Model for Mental Model Development

In this section, an adaptive network model is presented that simulates the development of mental models based on different theories on the formation of mental models. But first, for a better understanding of the adaptive network model addressed, the following familiar scenario is presented.

> Person A has rudimentary preconceptions about how to drive a car, the car's components, and their interactions. This mental model of driving which is incomplete and inaccurate, is activated as Person A enrolls for driving lessons. Person A's first driving lesson begins with instructor B initially demonstrating how to start a car and get it into a moving state. The initial observation of B improves the raw mental model that A has. As A attempts to learn, an iterative process of modifying actions and testing takes place. As this happens A learns new connections between different components leading to a more accurate and complete mental model. This process is further influenced by incorporating the incoming information from B who is in the instructor's seat helping A learn.

Temporal-causal network models as addressed in Treur (2016) can be represented by a conceptual representation and by a numerical representation. A conceptual representation involves representing in a declarative manner states and connections between them that represent the causal impacts of states on each other. The states are assumed to have activation levels that vary over time. In reality, not all causal relations are equally strong, so some notion of the strength of a connection is used. Furthermore, when more than one causal relation affects a state, some way to aggregate multiple causal impacts on a state is used. Therefore, for each state, a way to specify how multiple causal impacts on this state are aggregated is indicated by some combination function. For this aggregation, a number of standard combination functions are available as options and a number of desirable properties of such combination functions have been identified. Moreover, a notion of the speed of change of a state is used for the timing of the processes. The notions for *connectivity* (connection weights $\boldsymbol{\omega}_{X,Y}$), *aggregation* (combination functions $\mathbf{c}_Y(..)$) and *timing* (speed factors $\boldsymbol{\eta}_Y$) are the network characteristics defining a conceptual representation of a temporal-causal network model.

For adaptive networks the notion of *self-modeling network* (also called reified network) turns out useful, as has been worked out in (Treur 2020). A self-modeling network is obtained in the way that for each state Y of the base

network, for the adaptive ones among the network characteristics $\boldsymbol{\omega}_{X,Y}$, $\mathbf{c}_Y(..)$, $\boldsymbol{\eta}_Y$, additional network states (*self-model states*) are added to the network. For example, for adaptive connectivity characteristics, self-model states $\mathbf{RW}_{X,Y}$ are added representing adaptive connection weights $\boldsymbol{\omega}_{X,Y}$. To distinguish them from states of the base network, these self-model states are depicted at a separate level (see the blue plane in Figs. 3.1 and 3.2).

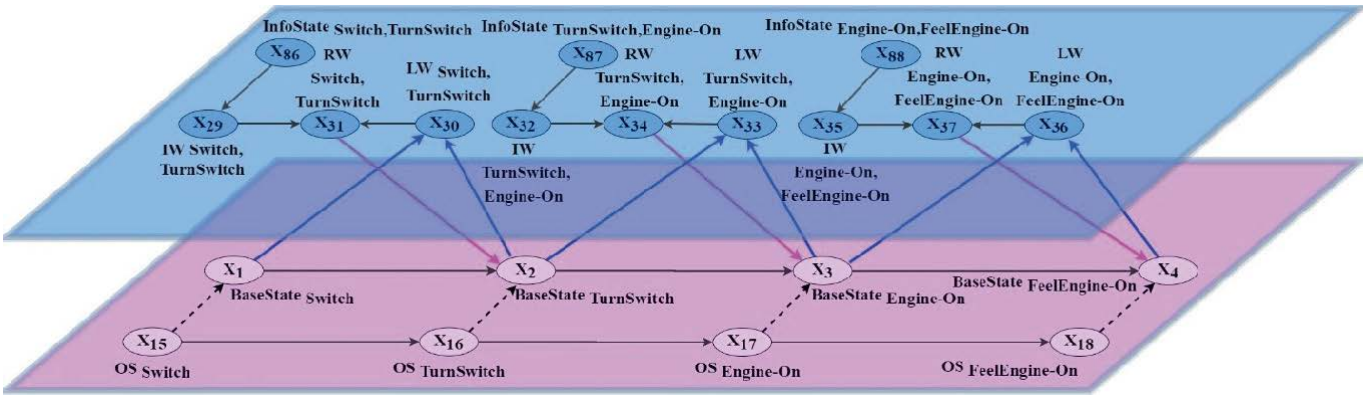


Fig. 3.1 Connectivity for part of the self-modeling network model

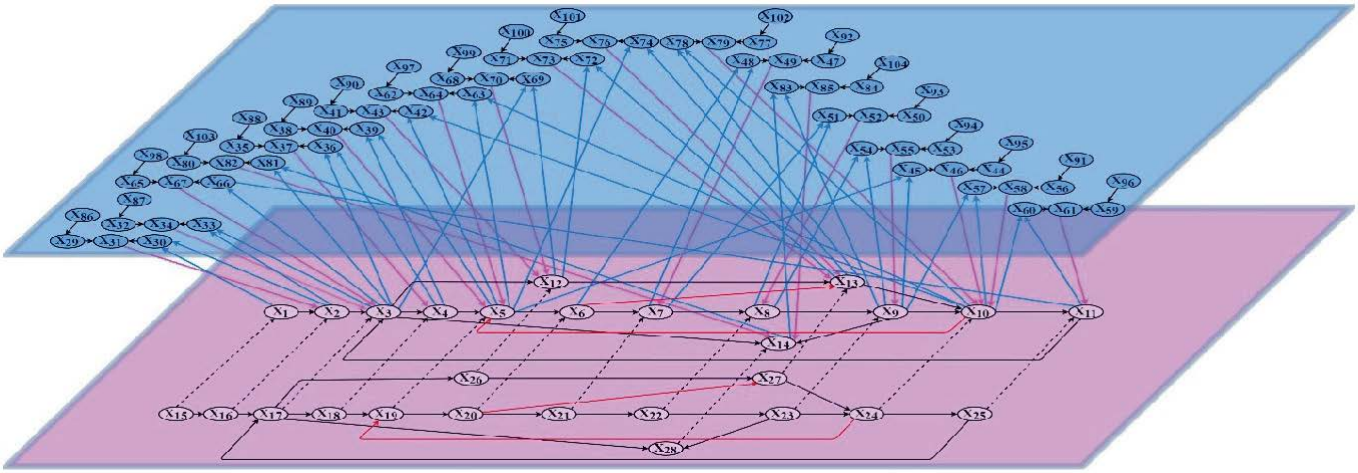


Fig. 3.2 Connectivity for the complete self-modeling network model

Mayer (1989) suggested that model-instruction provided to students can enhance the construction of mental models of the systems they are learning. Johnson-Laird (1989) identified three general elements of model-based learning and corresponding model-based instruction:

1. Everyday observations of the outside world (in the sense of observational learning and learning by imitation)
2. Other people´s explanations, especially in the classroom
3. The ability of an individual to construct mental models either from

components of world knowledge or from analogous models that the individual already possesses.

All three sources were incorporated into the current model.

In accordance with the distinction between base level and self-model level (also called reification level) briefly discussed above, the designed adaptive network model has these two levels. Each level is graphically depicted in 3D in one horizontal plane (see Figs. 3.1 and 3.2).

The lower plane contains the base network, whereas the higher plane represents the self-model states (in this case the InfoStates, **IW**-states, **LW**-states, and **RW**-states), all referring to connections between the Base States. Having these two levels within a temporal-causal network allows to manipulate both the states and the connections of the mental model, which, as discussed in Sect. 3.2, is required for the development process of a mental model. The structure of the two (interacting) levels allows to neatly distinguish the two types of processes (and their interaction): *using* the mental model by changing the Base States represented at the base level (*internal simulation* of the mental model) versus *adjusting* the mental model by changing the representations at the self-model level for its connections (*adaptation*, *learning* of the mental model). These different types of states are explained in Tables 3.1, 3.2 and 3.3; for a complete overview of them, see Sect. 3.6. Figure 3.1 depicts the connectivity for a part with just a small number of the states to make it more comprehensible. The connectivity for the complete network model consisting of 28 states at the base level and 3 * 28 = 84 states at the self-model level is shown in Fig. 3.2.

Table 3.1 Description of abbreviations used in the model

Abbreviation	Explanation
BS	Base states for the mental model of the learner
OS	Base states for the observations
LW	Self-model states for Hebbian learning of connections between Base States
IW	Self-model states for incoming information (instructions) from the instructor on connections between Base States
RW	Self-model states combining the learnt weights (**LW**) and information states(**IW**)
IS	Self-model states for the source of incoming information (instruction) from the instructor

Table 3.2 Explanation of the base level states in the model

State	Name	Explanation
X_1	$\mathbf{BS}_{\text{Switch}}$	Learner's Representation State for Switch
X_2	$\mathbf{BS}_{\text{TurnSwitch}}$	Learner's Representation State for TurnSwitch

State	Name	Explanation
X_3	**BS**$_{\text{Engine-On}}$	Learner's Representation State for Engine-On
X_4	**BS**$_{\text{FeelEngine-On}}$	Learner's Representation State for FeelEngine-On
X_5	**BS**$_{\text{PresClutch}}$	Learner's Representation State for PressClutch
X_6	**BS**$_{\text{Clutch-On}}$	Learner's Representation State for Clutch-On
X_7	**BS**$_{\text{Gearbox-Neutral}}$	Learner's Representation State for Gearbox-Neutral
X_8	**BS**$_{\text{PressGear 1}}$	Learner's Representation State for PressGear1
X_9	**BS**$_{\text{Gear1-On}}$	Learner's Representation State for Gear1-On
X_{10}	**BS**$_{\text{PressAccelerator}}$	Learner's Representation State for PressAccelerator
X_{11}	**BS**$_{\text{Accelerator-On}}$	Learner's Representation State for Accelerator-On
X_{12}	**BS**$_{\text{RevMeter-On}}$	Learner's Representation State for Rev-Meter-On
X_{13}	**BS**$_{\text{BiteState}}$	Learner's Representation State for BiteState
X_{14}	**BS**$_{\text{MovingState}}$	Learner's Representation State for MovingState
X_{15}	**OS**$_{\text{Switch}}$	Observation State for Switch
.....		
X_{28}	**OS**$_{\text{MovingState}}$	Observation State for MovingState

Table 3.3 Explanation of the self-model states in the part of the model depicted in Fig. 3.1; for an explanation of all states, see the Tables 3.4, 3.5 and 3.6 in Sect. 3.6

State	Name	Explanation
X_{29}	**IW**$_{\text{Switch, TurnSwitch}}$	Self-model state for Informed Weight for Switch → TurnSwitch
X_{30}	**LW**$_{\text{Switch, TurnSwitch}}$	Self-model state for Learnt Weight for Switch → TurnSwitch
X_{31}	**RW**$_{\text{Switch, TurnSwitch}}$	Self-model state for overall Representative Weight for Switch → TurnSwitch
X_{32}	**IW**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for Informed Weight for TurnSwitch → Engine-On
X_{33}	**LW**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for Learnt Weight for TurnSwitch → Engine-On
X_{34}	**RW**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for overall Weight for TurnSwitch → Engine-On
X_{35}	**IW**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for Informed Weight for Engine-On → FeelEngine-On
X_{36}	**LW**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for Learnt Weight for Engine-On → FeelEngine-On
X_{37}	**RW**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for overall Weight for Engine-On → FeelEngine-On
.....		

State Name		Explanation
X_{86}	**IS**$_{\text{Switch, TurnSwitch}}$	Self-model state for Information State for Switch → TurnSwitch
X_{87}	**IS**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for Information State for TurnSwitch → Engine-On
X_{88}	**IS**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for Information State for Engine-On → FeelEngine-On
.....		

3.3.1 Observational Learning

As mentioned earlier, observation or imitation of others is one of the sources that help the formation of mental models. In model-centered learning, trainees watch someone else perform a target behavior and then attempt to reenact it; e.g., Yi and Davis (2003). Action demonstration is a widely utilized method in teaching new motor tasks. This process of learning is referred to as observational motor learning. Empirical findings show that observational motor learning improves action perception and motor execution. Mirror Neurons are assumed to be responsible for the ability to learn by observing or imitating others as they help us understand the actions made by others; e.g., Rizzolatti and Craighero (2004). Mirror neurons are own preparation states that in addition are also activated by observations relating to these states of someone else. This means that the observation states have some connection to the mirroring states. In the model addressed here, this mirror neuron function of observational learning is modeled by having a direct connection from the Observation States (OS) to the Base States (BS); see Table 3.1 for these abbreviations. Here Base States represent the mental states of the model of the learner, whereas the Observation States represent the (by the learner) observed actions of someone (e.g., the instructor driving the car) demonstrating the process covered by the mental model.

So, this indeed gives the base states a similar function like mirror neurons have: they help us understand and learn the task at hand by improving the mental model based on observations. Table 3.2 represents all the states involved at the base level of the model. By the mirror neuron function of the Base States, the **OS**-states affect the Base States of the learner, which in turn amplifies the Hebbian learning that improves the learner's initial mental model, as discussed next. In Table 3.3 the self-model states (depicted in the blue plane) are explained.

3.3.2 Self-directed Learning

Self-guided learning occurs as a multi-step process of model-building and revision (Penner 2001). Johnson-Laird (1983) described this process as a procedure of 'fleshing out', i.e., that the learner continuously examines whether a model can be replaced with an alternative model or not. Buckley (2000)

describes how mental models are constructed. First, in the model formation part, existing knowledge structures are activated, then new information and demands of the tasks are incorporated to construct a mental model. If the constructed model is successful to interact with or reason about a phenomenon or accomplish a task (in the case considered here drive a car), then it is reinforced.

However, if the model is inconsistent or deficient with the task at hand, either the model is rejected and a new one is formed or the existing model is revised or elaborated. As explained in the scenario before, the learner builds on the initial rudimentary model by engaging in the task itself. The learner's initial model of driving the car evolves by the process of model revision and elaboration that involves building new connections and reinforcing the ones that work. For example, understanding the causal relation among different components while removing inaccurate connections. This process of self-learning is modeled by utilizing a Hebbian combination function for the **LW**-states for adaptive connection weights between any two base states. Learning causal relationships between parts of the mental model also involves understanding and modifying the mental representation of negative causal relationships they might have. In the current model, which is based on an old fashioned manual car, for example, there exists a negative causal relationship between Base State PressClutch and Base State PressAccelerator. This is modeled by using a different type of Hebbian combination function for handling negative weights for **LW**-states representing the learnt weights for the connection between these base states.

3.3.3 Learning from Instruction

It is understood that for a beginner learner, learning by discovery involves a great deal of trial and error (Seel 2006). Hence, along with self-learning, instructions from an expert are considered useful to build accurate and effective mental models. This notion is supported by scaffolded model-based learning in which a variety of supports like prompts, questions, hints, stories, conceptual models, visualizations are provided to assist the students' progress during learning tasks; e.g., Hogan and Pressley (1997). In the presented model, this is modeled by having a self-model state IW-state, each with a connection incoming information source (from the instructor) called Info State.

3.3.4 Integrating Self-Directed Learning and Learning from Instruction

Gibbons and Gray (2002) integrated both sources of learning by stating that instructions serve human learning processes under the control of the individual. Thus instructions do not cause learning but rather support it. The scaffolded model-based learning mentioned above supports this integration. This is

modeled in the current model by combining the weight values of the **LW**-states (self-directed learning states) with those of the **IW**-states (scaffolded model-based learning states) in an **RW**-state which serves to integrate these two specific learning results in an overall learning result. In using the mental model (by internal simulation), an **RW**-state value works as an (adaptive) weight for the connection between the two concerning base states.

The conceptual representation of a temporal–causal network model like the one mentioned above can be transformed in a systematic and automated manner into the numerical representation using a dedicated modeling environment resulting in difference and differential equations; see Treur (2020, Chaps. 2 and 9):

$$Y(t+\Delta t) = Y(t) + \boldsymbol{\eta}_Y \left[\mathbf{aggimpact}_Y(t) - Y(t)\right] \Delta t \tag{3.1}$$

or

$$\frac{dY(t)}{dt} = \boldsymbol{\eta}_Y \left[\mathbf{aggimpact}_Y(t) - Y(t)\right] \Delta t \tag{3.2}$$

where

$$\mathbf{aggimpact}_Y(t) = \mathbf{c}_Y\left(\boldsymbol{\omega}_{X_1,Y} X_1(t), \ldots, \boldsymbol{\omega}_{X_k,Y} X_k(t)\right) \tag{3.3}$$

In the model presented here, for the states, the following combination functions were used. The *Euclidean combination function* $\mathbf{eucl}_{n,\lambda}(V_1, ..., V_k)$ where n is the order (which can be any non-zero natural number, but also any positive real number), and λ the scaling factor is defined by:

$$\mathbf{eucl}_{n,\lambda}(V_1, \ldots, V_k) = \sqrt[n]{\frac{V_1^n + \cdots + V_k^n}{\lambda}} \tag{3.4}$$

The *Advanced logistic sum combination function* $\mathbf{alogistic}_{\sigma,\tau}(...)$ is used with steepness σ and threshold τ and is defined by:

$$\mathbf{alogistic}_{\sigma,\tau}(V_1, \ldots, V_k) = \left[\frac{1}{1+e^{-\sigma(V_1+\ldots+V_k-\tau)}} - \frac{1}{1+e^{\sigma\tau}}\right]\left(1+e^{-\sigma\tau}\right) \tag{3.5}$$

The *Hebbian learning combination function* $\mathbf{hebb}_{\mu}(..)$ for learning of the connection from state X to state Y is defined by

$$\mathbf{hebb}_{\mu}(V_1, V_2, W) = V_1 V_2 (1 - W) + \mu W \tag{3.6}$$

where μ is the persistence parameter, V_1 stands for state value $X(t)$, V_2 for $Y(t)$, and W for the learnt connection weight self-model state value $\mathbf{LW}_{X,Y}(t)$. Here the part $V_1 V_2 (1 - W)$ models the learning effect, which is based on maximal learning

for simultaneous high values V_1, V_2 of the two connected states; this is modeled by V_1V_2. Moreover, a decreasing learning effect is achieved if the value 1 for W is approached (so, if 1 – W becomes small), to keep the value of W within the interval [0, 1]; this is modeled by the factor 1 – W. The part $\boldsymbol{\mu}W$ models persistence. For the learning of negative weights, similarly, the criterion for maximal learning should be that V_1 has a high value while V_2 simultaneously has a low value; this is modeled by a high value for $V_1(1 - V_2)$. Moreover, for a negative weight, the learning effect should decrease when W approaches the value −1 (so $-1 - W = -(1 + W)$ becomes small). The part $\boldsymbol{\mu}W$ models persistence, which works the same for negative W. Based on the above, the function **hebbneg$_{\boldsymbol{\mu}}$(..)** for negative connection weights is derived from the formula for $\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W)$ by the following substitutions in the part $V_1V_2\,(1 - W)$ of the formula:

$$V_2 ==> 1 - V_2$$

$$1 - W ==> -(1 + W) \tag{3.7}$$

So, the combination function used for Hebbian learning of negative connection weights is:

$$\mathbf{hebbneg}_{\boldsymbol{\mu}}(V_1, V_2, W) = -V_1(1 - V_2)(1 + W) + \boldsymbol{\mu}W \tag{3.8}$$

Finally, the **stepmod** function is used for a repetitive input of a given time duration.

3.4 Example Simulations

The computational model was simulated using a dedicated software environment implemented in MATLAB described in Treur (2020, Chap. 9), to simulate and study the development of mental models; see Figs. 3.3, 3.4 and 3.5. For the simulation $\Delta t = 0.5$ was chosen, the total time 800. The speed factor for the Base States were set at 0.4, for **OS**-states at 0.05, for **IW**-states at 0.1, and for **LW**-states and **RW**-states at 0.4. The connection weights between the states and the other characteristics of the network model and the initial values are shown at https://www.researchgate.net/publication/342521927. All Base States have initial value 0. All **OS**-States have initial value 0, except the first **OS**-State X_{15} which has an initial value of 1. For all the **IW**- and **LW**-states the initial value was set at 0.1, while for the **RW**-states it's set at 0. Note that it has been set in such a way that only one of the **IW**-state or **LW**-state is not enough to get an **RW**-state at 1. A typical pattern is that first, based on an **IW**-state triggered by an Info State,

the **RW**-state gets a value somewhere in the middle of the 0–1 interval, and only after learning making the **LW**-state high, the **RW**-state value increases to 1.

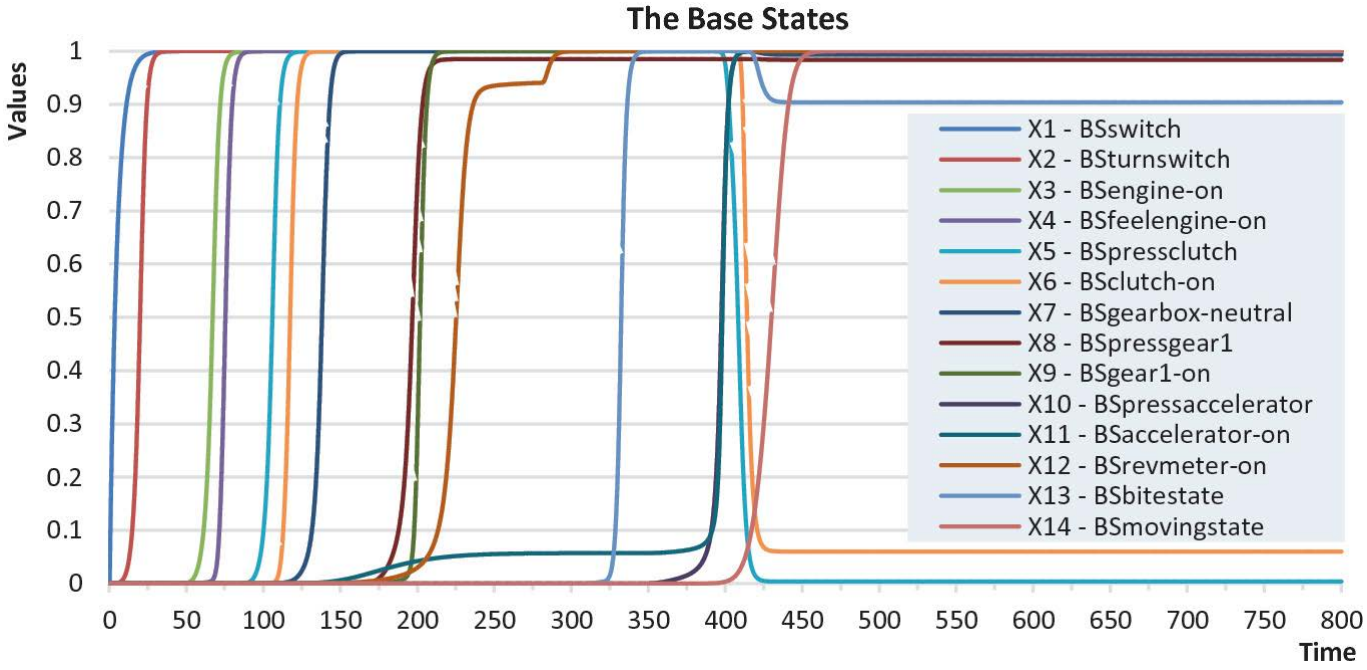


Fig. 3.3 Dynamics of the Base States X_1–X_{14} showing internal simulation of the mental model

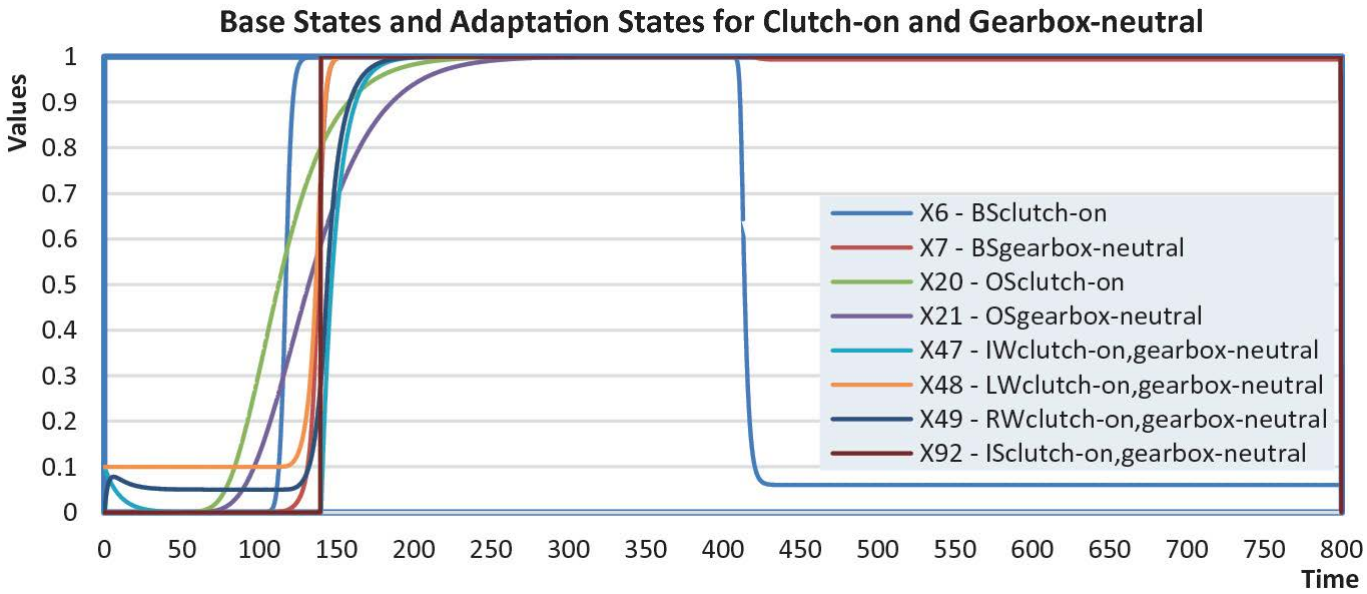


Fig. 3.4 Base States X_6 (Clutch-on) and X_7 (Gearbox-neutral) with impact from **OS**-states X_{20}, X_{21} and **LW**-state X_{48}, **RW**-state X_{49} and **IW**-state X_{47} with impact from **IS**-state X_{92}

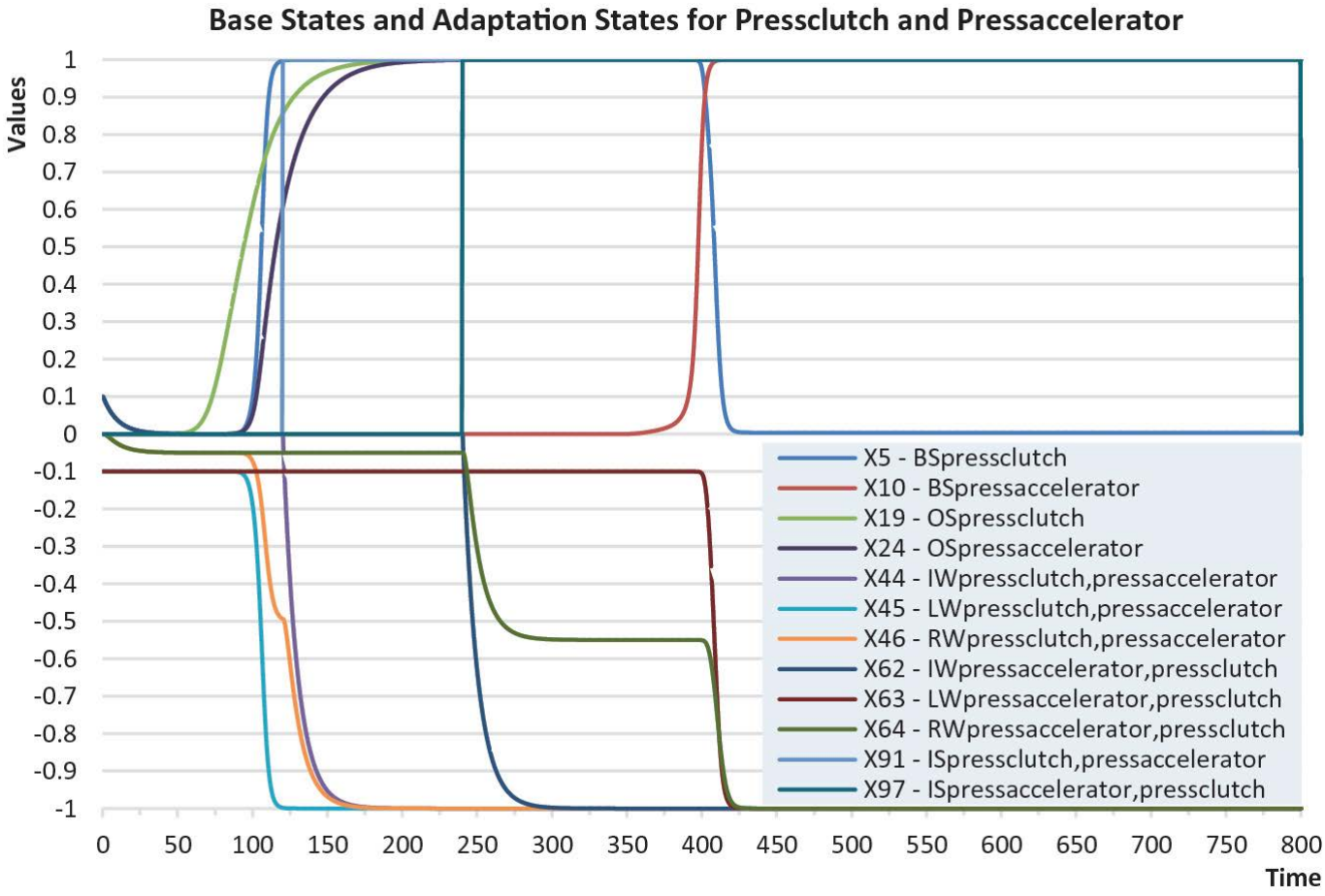


Fig. 3.5 Base States X_5 (Pressclutch) and X_{10} (Pressaccelerator) with impact from **OS**-states X_{19} and X_{24} and **LW**-state X_{63}, **RW**-state X_{64}, and **IW**-state X_{62} with impact from **IS**-states X_{91} and X_{97}

This reflects the assumption that by 'hearsay' from an instructor, no full knowledge is achieved, but still it is a good basis to complete the learning process to full knowledge by furthering one's own observational learning; see also the part on Integrating Self-Directed Learning and Learning from Instruction in Sect. 3.3. For simplification, the results indicated in Fig. 3.4 represent the connection between two base states X_6 and X_7. Here it can be seen that, together with the value of X_6 becoming 1 at time 110. The **OS**-state X_{21} affects the value of X_7 together with **RW**-state X_{49} which combines the weights of the related **LW**-and **IW**-states. The state X_7 reaches value 1 at time 190 by an S-curve. For the modeling of the negative loop for base states X_5 and X_{10} (Fig. 3.5) with the **hebbneg** combination function, the Info States for both have a negative incoming connection to the **IW**-states and a weight of 0 for connection to itself.

3.5 Discussion and Conclusion

In the present chapter, the development of mental models was studied based on literature; material for it was adopted from Bhalwankar and Treur (2021). First, a

deep literature study was conducted in identifying the processes and different theories involved from different disciplines. Then, a familiar scenario with respect to learning of device models was identified and chosen (learning to drive a car). Based on the literature, a conceptual representation of a self-modeling network model was designed. This model captures the development of mental models and as an illustration it was applied to learning to drive a car. Different theories exist which describe the development of mental models and are mentioned above. But, at the time the model proposed here was designed, as far as the authors know a formalized computational model for them was never designed, perhaps with an exception for (Van Ments et al. 2018) where a relatively simple adaptive God model was used as a mental model.

Based on the developed conceptual representation a formalized numerical representation and an implementation in the dedicated software environment developed in MATLAB were obtained. With this implemented adaptive network model, the development of mental models while learning to drive a car based on literature was simulated and shown to work as expected from the literature.

The simulation of the development of mental models shows interesting results. While the model does represent how mental models are formed based on literature there are still things to improve. For example, more insights from neuroscience of how such mental models are formed can be incorporated to make it more valid and representative of the processes in human beings. The process of formalizing such a complex process is iterative, where further adaptations from literature study, development of specific applications and their simulations can be made to match the phenomenon of mental models in humans as much as possible, while preserving the genericity of the computational model.

3.6 Explanation of All States of the Model

In this section, in three tables all states of the network model are explained. In Table 3.4 all base states are explained, in Table 3.5 all self-model **IW**-, **LW**-, and **RW**-states, and in Table 3.6 all self-model **IS**-states. An Appendix with a detailed description of the entire adaptive network model can be found as Linked Data at https://www.researchgate.net/publication/342521927.

Table 3.4 Explanation of all basestates

State		Explanation
X_1	$\mathbf{BS}_{\text{Switch}}$	Learner's Representation State for Switch
X_2	$\mathbf{BS}_{\text{TurnSwitch}}$	Learner's Representation State for TurnSwitch
X_3	$\mathbf{BS}_{\text{Engine-0n}}$	Learner's Representation State for Engine-On

State		Explanation
X_4	**BS**$_{\text{FeelEngine-On}}$	Learner's Representation State for FeelEngine-On
X_5	**BS**$_{\text{PresClutch}}$	Learner's Representation State for PressClutch
X_6	**BS**e$_{\text{Clutch-On}}$	Learner's Representation State for Clutch-On
X_7	**BS**$_{\text{Gearbox-Neutral}}$	Learner's Representation State for Gearbox-Neutral
X_8	**BS**$_{\text{PressGear 1}}$	Learner's Representation State for PressGear1
X_9	**BS**$_{\text{Gear1-On}}$	Learner's Representation State for Gear1-On
X_{10}	**BS**$_{\text{PressAccelerator}}$	Learner's Representation State for PressAccelerator
X_{11}	**BS**$_{\text{Accelerator-On}}$	Learner's Representation State for Accelerator-On
X_{12}	**BS**$_{\text{RevMeter-On}}$	Learner's Representation State for Rev-Meter-On
X_{13}	**BS**$_{\text{BiteState}}$	Learner's Representation State for Bite State
X_{14}	**BS**$_{\text{MovingState}}$	Learner's Representation State for Moving State
X_{15}	**OS**$_{\text{Switch}}$	Observation State for Switch
X_{16}	**OS**$_{\text{TurnSwitch}}$	Observation State for Turn Switch
X_{17}	**OS**$_{\text{Engine-0n}}$	Observation State for Engine-On
X_{18}	**OS**$_{\text{FeelEngine-On}}$	Observation State for Feel Engine-On
X_{19}	**OS**$_{\text{PressClutch}}$	Observation State for Press Clutch
X_{20}	**OS**$_{\text{Clutch-On}}$	Observation State for Clutch-On
X_{21}	**OS**$_{\text{Gearbox-Neutral}}$	Observation State for Gearbox-Neutral
X_{22}	**OS**$_{\text{PressGear 1}}$	Observation State for PressGear1
X_{23}	**OS**$_{\text{Gear1}}$	Observation State for Gear1-On
X_{24}	**OS**$_{\text{PressAccelerator}}$	Observation State for Press Accelerator
X_{25}	**OS**$_{\text{Accelerator-On}}$	Observation State for Accelerator-On
X_{26}	**OS**$_{\text{RevMeter-On}}$	Observation State for Rev-Meter-On
X_{27}	**OS**$_{\text{BiteState}}$	Observation State for Bite State
X_{28}	**OS**$_{\text{MovingState}}$	Observation State for Moving State

Table 3.5 Explanation of all self-model **IW**-, **LW**- and **RW**-states

States		Explanation
X_{29}	**IW**$_{\text{Switch, TurnSwitch}}$	Self-model state for Informed Connection Weight Switch → TurnSwitch
X_{30}	**LW**$_{\text{Switch, TurnSwitch}}$	Self-model state for Learnt Connection Weight Switch → TurnSwitch
X_{31}	**RW**$_{\text{Switch, TurnSwitch}}$	Self-model state for overall Connection Weight Switch → TurnSwitch

States		Explanation
X_{32}	**IW**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for Informed Connection Weight TurnSwitch → Engine-On
X_{33}	**LW**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for Learnt Connection Weight TurnSwitch → Engine-On
X_{34}	**RW**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for overall Connection Weight TurnSwitch → Engine-On
X_{35}	**IW**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for Informed Connection Weight Engine-On → FeelEngine-On
X_{36}	**LW**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for Learnt Connection Weight Engine-On → FeelEngine-On
X_{37}	**RW**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for overall Connection Weight Engine-On → FeelEngine-On
X_{38}	**IW**$_{\text{FeelEngine-On, PressClutch}}$	Self-model state for Informed Connection Weight FeelEngine-On → PressClutch
X_{39}	**LW**$_{\text{FeelEngine-On, PressClutch}}$	Self-model state for Learnt Connection Weight FeelEngine-On → PressClutch
X_{40}	**RW**$_{\text{FeelEngine-On, PressClutch}}$	Self-model state for overall Connection Weight FeelEngine-On → PressClutch
X_{41}	**IW**$_{\text{PressClutch, Clutch-On}}$	Self-model state for Informed Connection Weight PressClutch → Clutch-On
X_{42}	**LW**$_{\text{PressClutch, Clutch-On}}$	Self-model state for Learnt Connection Weight PressClutch → Clutch-On
X_{43}	**RW**$_{\text{PressClutch, Clutch-On}}$	Self-model state for overall Connection Weight PressClutch → Clutch-On
X_{44}	**IW**$_{\text{PressClutch, PressAccelerator}}$	Self-model state for Informed Connection Weight PressClutch → PressAccelerator
X_{45}	**LW**$_{\text{PressClutch, PressAccelerator}}$	Self-model state for Learnt Connection Weight PressClutch → PressAccelerator
X_{46}	**RW**$_{\text{PressClutch, PressAccelerator}}$	Self-model state for overall Connection Weight PressClutch → PressAccelerator
X_{47}	**IW**$_{\text{Clutch-On, Gearbox-Neutral}}$	Self-model state for Informed Connection Weight Clutch-On → Gearbox-Neutral
X_{48}	**LW**$_{\text{Clutch-On, Gearbox-Neutral}}$	Self-model state for Learnt Connection Weight Clutch-On → Gearbox-Neutral
X_{49}	**RW**$_{\text{Clutch-On, Gearbox-Neutral}}$	Self-model state for overall Connection Weight Clutch-On → Gearbox-Neutral
X_{50}	**IW**$_{\text{Gearbox-Neutral, PressGear1}}$	Self-model state for Informed Connection Weight Gearbox-Neutral → PressGear1
X_{51}	**LW**$_{\text{Gearbox-Neutral, PressGear1}}$	Self-model state for Learnt Connection Weight Gearbox-Neutral → PressGear1

States		Explanation
X_{52}	**RW**$_{\text{Gearbox-Neutral, PressGear1}}$	Self-model state for overall Connection Weight Gearbox-Neutral → PressGear1
X_{53}	**IW**$_{\text{PressGear1, Gear1-On}}$	Self-model state for Informed Connection Weight PressGear1 → Gear1-On
X_{54}	**LW**$_{\text{PressGear1, Gear1-On}}$	Self-model state for Learnt Connection Weight PressGear1 → Gear1-On
X_{55}	**RW**$_{\text{PressGear1, Gear1-On}}$	Self-model state for overall Connection Weight PressGear1 → Gear1-On
X_{56}	**IW**$_{\text{Gear1-On, PressAccelerator}}$	Self-model state for Informed Connection Weight Gear1-On → PressAccelerator
X_{57}	**LW**$_{\text{Gear1-On, PressAccelerator}}$	Self-model state for Learnt Connection Weight Gear1-On → PressAccelerator
X_{58}	**RW**$_{\text{Gear1-On, PressAccelerator}}$	Self-model state for overall Connection Weight Gear1 → On, PressAccelerator
X_{59}	**IW**$_{\text{PressAccelerator, Accelerator-On}}$	Self-model state for Informed Connection Weight PressAccelerator → Accelerator-On
X_{60}	**LW**$_{\text{PressAccelerator, Accelerator-On}}$	Self-model state for Learnt Connection Weight PressAccelerator → Accelerator-On
X_{61}	**RW**$_{\text{PressAccelerator, Accelerator-On}}$	Self-model state for overall Connection Weight PressAccelerator → Accelerator-On
X_{62}	**IW**$_{\text{PressAccelerator, PressingClutch}}$	Self-model state for Informed Connection Weight PressAccelerator → PressingClutch
X_{63}	**LW**$_{\text{PressAccelerator, PressingClutch}}$	Self-model state for Learnt Connection Weight PressAccelerator → PressingClutch
X_{64}	**RW**$_{\text{PressAccelerator, PressingClutch}}$	Self-model state for overall Connection Weight PressAccelerator → PressingClutch
X_{65}	**IW**$_{\text{Accelerator-On, Engine-On}}$	Self-model state for Informed Connection Weight Accelerator-On → Engine-On
X_{66}	**LW**$_{\text{Accelerator-On, Engine-On}}$	Self-model state for Learnt Connection Weight Accelerator-On → Engine-On
X_{67}	**RW**$_{\text{Accelerator-On, Engine-On}}$	Self-model state for overall Connection Weight Accelerator-On → Engine-On
X_{68}	**IW**$_{\text{Engine-On, Rev-Meter-On}}$	Self-model state for Informed Connection Weight Engine-On → Rev-Meter-On
X_{69}	**LW**$_{\text{Engine-On, Rev-Meter-On}}$	Self-model state for Learnt Connection Weight Engine-On → Rev-Meter-On
X_{70}	**RW**$_{\text{Engine-On, Rev-Meter-On}}$	Self-model state for overall Connection Weight Engine-On → Rev-Meter-On
X_{71}	**IW**$_{\text{RevMeter-On, BiteState}}$	Self-model state for Informed Connection Weight RevMeter-On → BiteState

States		Explanation
X_{72}	**LW**$_{\text{RevMeter-On, BiteState}}$	Self-model state for Learnt Connection Weight RevMeter-On → BiteState
X_{73}	**RW**$_{\text{RevMeter-On, BiteState}}$	Self-model state for overall Connection Weight RevMeter-On → BiteState
X_{74}	**IW**$_{\text{Clutch-On, BiteState}}$	Self-model state for Informed Connection Weight Clutch-On → BiteState
X_{75}	**LW**$_{\text{Clutch-On, BiteState}}$	Self-model state for Learnt Connection Weight Clutch-On → BiteState
X_{76}	**RW**$_{\text{Clutch-On, BiteState}}$	Self-model state for overall Connection Weight Clutch-On → BiteState
X_{77}	**IW**$_{\text{BiteState, PressAccelerator}}$	Self-model state for Informed Connection Weight BiteState → PressAccelerator
X_{78}	**LW**$_{\text{BiteState, PressAccelerator}}$	Self-model state for Learnt Connection Weight BiteState → PressAccelerator
X_{79}	**RW**$_{\text{BiteState, PressAccelerator}}$	Self-model state for overall Connection Weight BiteState → PressAccelerator
X_{80}	**IW**$_{\text{Engine-On, MovingState}}$	Self-model state for Informed Connection Weight Engine-On → MovingState
X_{81}	**LW**$_{\text{Engine-On, MovingState}}$	Self-model state for Learnt Connection Weight Engine-On → MovingState
X_{82}	**RW**$_{\text{Engine-On, MovingState}}$	Self-model state for overall Connection Weight Engine-On → MovingState
X_{83}	**IW**$_{\text{Gear1-On, MovingState}}$	Self-model state for Informed Connection Weight Gear1-On → MovingState
X_{84}	**LW**$_{\text{Gear1-On, MovingState}}$	Self-model state for Learnt Connection Weight Gear1-On → MovingState
X_{85}	**RW**$_{\text{Gear1-On, MovingState}}$	Self-model state for overall Connection Weight Gear1-On → MovingState

Table 3.6 Explanation of all self-model **IS**-states

States		Explanation
X_{86}	**IS**$_{\text{Switch, TurnSwitch}}$	Self-model state for Information State for Switch → TurnSwitch
X_{87}	**IS**$_{\text{TurnSwitch, Engine-On}}$	Self-model state for Information State for TurnSwitch → Engine-On
X_{88}	**IS**$_{\text{Engine-On, FeelEngine-On}}$	Self-model state for Information State for Engine-On → FeelEngine-On
X_{89}	**IS**$_{\text{FeelEngineOn, PressClutch}}$	Self-model state for Information State for FeelEngineOn → PressClutch
X_{90}	**IS**$_{\text{PressClutch, Clutch-On}}$	Self-model state for Information State for PressClutch → Clutch-On
X_{91}	**IS**$_{\text{PressClutch, PressAccelerator}}$	Self-model state for Information State for PressClutch → PressAccelerator
X_{92}	**IS**$_{\text{Clutch-On, Gear-BoxNeutral}}$	Self-model state for Information State for Clutch-On → Gear-BoxNeutral

States		Explanation
X_{93}	**IS**$_{GearBox\text{-}Neutral, PressGear1}$	Self-model state for Information State for GearBox-Neutral, PressGear1
X_{94}	**IS**$_{PressGear1, Gear1\text{-}On}$	Self-model state for Information State for PressGear1 → Gear1-On
X_{95}	**IS**$_{Gear1\text{-}On, PressAccelerator}$	Self-model state for Information State for Gear1-On → PressAccelerator
X_{96}	**IS**$_{PressAccelerator, Accelerator\text{-}On}$	Self-model state for Information State for PressAccelerator → Accelerator-On
X_{97}	**IS**$_{PressAccelerator, PressClutch}$	Self-model state for Information State for PressAccelerator → PressClutch
X_{98}	**IS**$_{Accelerator\text{-}On, Engine\text{-}On}$	Self-model state for Information State for Accelerator-On → Engine-On
X_{99}	**IS**$_{Engine\text{-}On, Rev\text{-}Meter\text{-}On}$	Self-model state for Information State for Engine-On → Rev-Meter-On
X_{100}	**IS**$_{RevMeter\text{-}On, BiteState}$	Self-model state for Information State for RevMeter-On → BiteState
X_{101}	**IS**$_{Clutch\text{-}On, BiteState}$	Self-model state for Information State for Clutch-On → BiteState
X_{102}	**IS**$_{BiteState, PressAccelerator}$	Self-model state for Information State for BiteState → PressAccelerator
X_{103}	**IS**$_{Engine\text{-}On, MovingState}$	Self-model state for Information State for Engine-On → MovingState
X_{104}	**IS**$_{Gear\text{-}On1, MovingState}$	Self-model state for Information State for Gear-On1 → MovingState

References

Bandura, A., Walters, R.H.: Social learning theory, vol. 1. Prentice-Hall, Englewood Cliffs, NJ

Barquero, B.: La representación de estados mentales en la comprensión de textos desde el enfoque teórico de los modelos mentales. Ph.D. thesis, University of Autónoma de Madrid (1995)

Benbassat, J.: Role modeling in medical education: the importance of a reflective imitation. Acad. Med. **89**(4), 550–554 (2014)
[Crossref]

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 Comput. Sci. **90**(4), 90–101. Elsevier (2021)

Buckley, B.C.: Interactive multimedia and model-based learning in biology. Int. J. Sci. Educ. **22**(9), 895–935 (2000)
[Crossref]

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

De Kleer, J., Brown, J.: Assumptions and ambiguities in mechanistic mental models. Gentner, D., Stevens, A. (eds.) Mental Models, pp. 155–190. Lawrence ErlbaumAssociates, Hillsdale, NJ (1983)

Devi, R., Tiberghien, A., Baker, M.J., Brna, P.: Modelling Students' construction of energy models in physics. Instr.

Sci. **24**, 259–293 (1996)
[Crossref]

Doll, B.B., Simon, D.A., Daw, N.D.: The ubiquity of model-based reinforcement learning. Curr. Opin. Neurobiol. **22**, 1075–1081 (2012)
[Crossref]

Ellison, C.G., Bradshaw, M., Kuyel, N., Marcum, J.P.: Attachment to God, stressful life events, and changes in psychological distress. Rev. Relig. Res. **53**(4), 493–511 (2012)
[Crossref]

Fein, R.M., Olson, G.M., Olson, J.S.: A mental model can help with learning to operate a complex device. In: The Interact'93 and Chi'93 Conferences Companion on Human Factors in Computing Systems, pp. 157–158. ACM Press, New York (1993)

Gentner, D., Stevens, A.L.: Mental Models. Erlbaum, Hillsdale NJ (1983)

Gibbons, J., Gray, M.: An integrated and experience-based approach to social work education: the Newcastle model. Soc. Work. Educ. **21**(5), 529–549 (2002)
[Crossref]

Greca, I.M., Moreira, M.A.: Mental models, conceptual models, and modelling. Int. J. Sci. Educ. **22**(1), 1–11 (2000)
[Crossref]

Halloun, I.: Schematic modelling for meaningful learning of physics. J. Res. Sci. Teach. **33**, 1019–1041 (1996)
[Crossref]

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

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

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)
[Crossref]

Johnson-Laird, P.N.: Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness (No. 6). Harvard University Press (1983)

Johnson-Laird, P.N.: Mental models. In: Posner, M.I. (ed.) Foundations of Cognitive Science, pp. 469–499. The MIT Press (1989)

Kieras, D.E., Bovair, S.: The role of a mental model in learning to operate a device. Cogn. Sci. **8**(3), 255–273 (1984)
[Crossref]

Kim, D.: The link between individual and organizational learning. In: Starkey, K., Tempest, S., Mckinlay, A. (eds.) How Organizations Learn, 2nd edn., pp. 29–50. Thomson Learning, London (2004)

Mayer, R.E.: Models for understanding. Rev. Educ. Res. **59**(1), 43–64 (1989)
[Crossref]

Moray, N.: The psychodynamics of human-machine interaction. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics, vol. 4, pp. 225–235. Ashgate (1999)

Nersessian, N.: Should physicists preach what they practice? Sci. Educ. **4**, 203–226 (1995)
[Crossref]

Olson, J.R.: The what and why of mental models in human computer interaction. In: Booth, P.A. (ed.), Mental Models in Everyday Activities, pp. 132–146. Robinson College (1992)

Penner, D.E.: Cognition, computers, and synthetic science: building knowledge and meaning through modeling. Rev. Res. Educ. **25**, 1–35 (2001)

Rizzolatti, G., Craighero, L.: The mirror-neuron system. Annu. Rev. Neurosci. **27**, 169–192 (2004)
[Crossref]

Rouse, W.B., Morris, N.M.: On looking into the black box: prospects and limits in the search for mental models. Psychol. Bull. **100**(3), 349–363 (1986)
[Crossref]

Rumelhart, D.E., Smolensky, P., McClelland, J.L., Hinton, G.E.: Schemata and sequential thought processes in PDP models. In: J.L. McClelland, D.E. Rumelhart, The PDP Research Group (eds.) Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Psychological and Biological Models, vol. 2, pp. 7–57. Cambridge, MIT Press, MA (1986)

Seel, N.M.: World Knowledge and Mental Models (in German). Hogrefe, Göttingen (1991)

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

Senge, P.: The Fifth Discipline: The Art and Practice of the Learning Organization. Doubleday, New York NY (1990)

Sutton, C.: The scientific model as a form of speech. Res. Sci. Educ. Europe **2**, 143–152 (1996)

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions. Springer, Cham, Switzerland (2016)
[Crossref]

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

Van Gog, T., Paas, F., Marcus, N., Ayres, P., Sweller, J.: The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. Educ. Psychol. Rev. **21**(1), 21–30 (2009)
[Crossref]

Van Ments, L., Roelofsma, P., Treur, J.: Modelling the effect of religion on human empathy based on an adaptive temporal–causal network model. Comput. Soc. Netw. **5**(1), Article 1 (2018)

Yi, M.Y., Davis, F.D.: Developing and validating an observational learning model of computer software training and skill acquisition. Inf. Syst. Res. **14**(2), 146–169 (2003)
[Crossref]

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_4

4. Controlling Your Mental Models: Using Metacognition to Control Use and Adaptation for Multiple Mental Models

Jan Treur[1] ✉
(1) Social AI Group, Department of Computer Science, Vrije Universiteit, Amsterdam, Netherlands

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
Learning processes can be described by adaptive mental (or neural) network models. If metacognition is used to regulate learning, the adaptation of the mental network becomes itself adaptive as well: second-order adaptation. In this chapter, a second-order adaptive mental network model is introduced for metacognitive regulation of learning processes. The focus is on the role of multiple internal mental models, in particular, the case of visualisation to support learning of numerical or symbolic skills. The second-order adaptive network model is illustrated by a case scenario for the role of visualisation to support learning multiplication at the primary school.

Keywords Metacognition – Control – Mental model – Multiple representation

4.1 Introduction

Metacognition (Darling-Hammond et al. 2008; Shannon 2008; Mahdavi 2014; Flavell 1979; Koriat 2007; Pintrich 2000) is a form of cognition about cognition. In (Koriat 2007) it is described as what people know about their own cognitive processes and how they put that knowledge to use in regulating their cognitive processing and behavior. A sometimes used closely related term is self-regulation and when the cognitive processes addressed by metacognition

concern learning, the term self-regulated learning is used. For example, in (Pintrich 2000), self-regulated learning is described as an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and context.

In learning, often different mental models play a role; e.g., (Gentner and Stevens 1983; Greca and Moreira 2000; Skemp 1971; Seel 2006). A specific case where the role of metacognition in learning processes is considered within educational science, is the use of multiple mental models such as in visualisation to support learning of more abstract (numerical or symbolic) skills; e.g., (Bruner 1966; Du Plooy 2016). An important metacognitive control decision in this context is whether or not and when to switch from one mental model to another one. In the educational science literature, much more can be found on this case, particularly for learning arithmetic or algebraic skills in primary or secondary schools supported by visualisation; see also (Bruner 1977; Bidwell 1972; Day and Hurell 2015; Freudenthal 1973; Freudenthal 1986; Koedinger and Terao 2002; Larbi and Mavis 2016; Lovitt et al. 1984; Renkema 2019; Roberts 1989).

From a network-oriented modeling perspective, learning is usually described by adaptive mental (or neural) network models, where some of the network characteristics such as connection weights or excitability thresholds change over time. If, in addition, metacognition is used to regulate or control the learning, this implies that the adaptation (by learning) of the mental network is itself adaptive as well, which is called second-order adaptation. Thus, a network model for such processes has to address such complex structures and behaviour. In the current chapter, using the modeling approach for higher-order adaptive networks from (Treur 2018,2020a,b), a second-order adaptive mental network model is introduced for metacognitive regulation of such learning processes. Here, the focus is on the role of multiple mental models in case of visualisation to support learning of more abstract (numerical or symbolic) skills. The adaptive network model is illustrated for a case study on the role of visualisation to support learning multiplication at the primary school as described, for example, in (Bruner 1966; Day and Hurrell 2015; Du Plooy 2016; Freudenthal 1973,1986; Rivera, 2011).

In this chapter, first in Sect. 4.2 more background knowledge is discussed on metacognition and the role of visualisation in learning processes. In Sect. 4.3 the network-oriented modeling approach used is briefly explained. Next, in Sect. 4.4 the introduced second-order adaptive network model is described in some detail. In Sect. 4.5, it is shown how this model was used to perform simulations for the illustrative example scenario. Finally, Sect. 4.6 is a discussion.

4.2 Metacognition and Multiple Mental Models

Literature on metacognition, sometimes also called self-regulation, can be found, for example in (Darling-Hammond et al. 2008; Shannon 2008; Mahdavi 2014; Flavell 1979; Koriat 2007; Pintrich 2000). The focus is here on the role of metacognition in learning. For example, in (Pintrich 2000, pp. 452–453) the following assumptions for self-regulated learning are described:

- It is a process whereby learners set goals for their learning and monitor and control their cognition, motivation, and behavior guided by these goals.
- Learners actively construct their own meanings, goals, and strategies.
- Learners can monitor, control, and regulate certain aspects of their own cognition, motivation, and behavior, and some elements of their environment.
- Some type of criterion or standard is used to assess whether the process should continue as is or if some type of change is necessary.
- Self-regulatory activities are mediators between personal and contextual characteristics and actual performance.

In line with these assumptions, in (Pintrich 2000, pp. 453–461, Table 1, p. 454), the following phases for self-regulation are described:

Table 4.1 The states in the adaptive network model

X_1	$\mathbf{N}_1$	Base state for number a
X_2	$\mathbf{N}_2$	Base state for number b
X_3	$\mathbf{N}_3$	Base state for number c
X_4	$\mathbf{S}_{23}$	Base state for number $b + c$
X_5	$\mathbf{P}_{12}$	Base state for number $a*b$
X_6	$\mathbf{P}_{13}$	Base state for number $a*c$
X_7	$\mathbf{PS}_{123}$	Base state for number $a*(b + c)$
X_8	$\mathbf{SP}_{1213}$	Base state for number $a*b + a*c$
X_9	$\mathbf{RD}_{vert}$	Vertical dimension of the rectangles
X_{10}	$\mathbf{RD}_{hor1}$	Horizontal dimension of rectangle 1
X_{11}	$\mathbf{RD}_{hor2}$	Horizontal dimension of rectangle 2
X_{12}	$\mathbf{RD}_{hor3}$	Horizontal dimension of rectangle 3
X_{13}	$\mathbf{RA}_1$	Area of rectangle 1
X_{14}	$\mathbf{RA}_2$	Area of rectangle 2
X_{15}	$\mathbf{RA}_3$	Area of rectangle 3
X_{16}	$\mathbf{RA}_{12}$	Area of rectangles 1 and 2 together
X_{17}	$\mathbf{W}_{P112}$	Representation state for the weight of the connection from $\mathbf{N}_1$ to $\mathbf{P}_{12}$
X_{18}	$\mathbf{W}_{P212}$	Representation state for the weight of the connection from $\mathbf{N}_2$ to $\mathbf{P}_{12}$
X_{19}	$\mathbf{W}_{P113}$	Representation state for the weight of the connection from $\mathbf{N}_1$ to $\mathbf{P}_{13}$
X_{20}	$\mathbf{W}_{P313}$	Representation state for the weight of the connection from $\mathbf{N}_3$ to $\mathbf{P}_{13}$
X_{21}	$\mathbf{W}_{SP121213}$	Representation state for the weight of the connection from $\mathbf{P}_{12}$ to $\mathbf{SP}_{1213}$
X_{22}	$\mathbf{W}_{SP131213}$	Representation state for the weight of the connection from $\mathbf{P}_{13}$ to $\mathbf{SP}_{1213}$
X_{23}	$\mathbf{RW}_P$	Mental representation state concerning the weights of the connections to $\mathbf{P}_{12}$ and $\mathbf{P}_{13}$
X_{24}	$\mathbf{RW}_{SP}$	Mental representation state concerning the weights of the connections to $\mathbf{SP}_{1213}$
X_{25}	$\mathbf{WRD}_{vert}$	Representation state used for execution of control decision $\mathbf{CWRD}_{vert}$, representing the weight of the connection from $\mathbf{N}_1$ to $\mathbf{RD}_{vert}$
X_{26}	$\mathbf{WRD}_{hor1}$	Representation state used for execution of control decision $\mathbf{CWRD}_{hor1}$, representing the weight of the connection from $\mathbf{N}_2$ to $\mathbf{RD}_{hor1}$
X_{27}	$\mathbf{WRD}_{hor2}$	Representation state used for execution of control decision $\mathbf{CWRD}_{hor2}$, representing the weight of the connection from $\mathbf{N}_3$ to $\mathbf{RD}_{hor2}$
X_{28}	$\mathbf{RS}_{num}$	Representation of the self-model for the own numerical skills
X_{29}	$\mathbf{RS}_{geo}$	Representation of the self-model for the own geometric skills
X_{30}	$\mathbf{CWRD}_{vert}$	Control state for the switch to the geometric mental model: representation of the weight of the connection from $\mathbf{RW}_{PSP}$ to $\mathbf{WRD}_{vert}$
X_{31}	$\mathbf{CWRD}_{hor1}$	Control state for the switch to the geometric mental model: representation of the weight of the connection from $\mathbf{RW}_{PSP}$ to $\mathbf{WRD}_{hor1}$
X_{32}	$\mathbf{CWRD}_{hor2}$	Control state for the switch to the geometric mental model: representation of the weight of the connection from $\mathbf{RW}_{PSP}$ to $\mathbf{WRD}_{hor2}$

- Cognitive planning and activation
- Cognitive monitoring
- Cognitive control and regulation
- Cognitive reaction and reflection

In (Koriat 2007, p. 290), metacognition is described by what people know about cognition and in particular their own cognitive processes, and how they use that in regulating their cognitive processes and behavior. Assumptions mentioned there are:

- Self-controlled cognitive processes have measurable effects on behavior (Koriat 2007), pp. 292–293
- Feelings, such as the feeling of knowing are part of monitoring, and exert a causal role on the control of cognitive processing (Koriat 2007), p. 293, p. 314–315
- There is a causal relation from monitoring to control (Koriat 2007), p. 315

So, in both descriptions of Pintrich (2000) and Koriat (2007) on metacognition (as well as in most other literature on metacognition), monitoring and control of the own cognitive processes are central concepts (where Koriat also emphasizes the feeling or experiencing that comes together with monitoring). These processes work through a causal cycle where the own cognitive processes affect the metacognitive monitoring, this monitoring in turn affects the metacognitive control, and this control affects the own cognitive processes. This causal cycle will indeed be incorporated in the adaptive network model introduced in Sect. 4. Note that metacognitive monitoring is usually based on forming and maintaining a self-model describing a (subjective) estimation of some relevant aspects of the own cognitive processes.

In the area of learning using multiple mental models (Gentner and Stevens 1983; Greca and Moreira 2000; Skemp 1971; Seel 2006), metacognition plays an important role for the decisions about when to switch from one mental model to another one. In particular, this takes place when learning numerical or symbolic skills in arithmetic or mathematics is supported by visualisations; e.g., see (Bruner 1966,1977; Bidwell 1972; Day and Hurell 2015; Du Plooy 2016; Freudenthal 1973,1986; Koedinger and Terao 2002; Larbi and Mavis 2016; Lovitt et al. 1984; Renkema 2019; Roberts 1989). Here, when at some point during working with a numerical or symbolic mental model, a learner monitors that the cognitive processes get stuck, the control decision can be made by the learner to switch to working with a mental model based on visualisation, after which the outcomes can be fed back to the numerical or symbolic mental model. Within the literature in educational science as mentioned above, it is extensively described how such a detour via a visualisation can support the learning of numerical or symbolic skills. This type of use of metacognition for using multiple mental models is the main focus in the current chapter.

4.3 Higher-Order Adaptive Network Models

In this section, the network-oriented modeling approach used is briefly introduced. Following (Treur 2016,2020b), a temporal-causal network model is characterised by (here X and Y denote nodes of the network, also called states):

- *Connectivity characteristics*

Connections from a state *X* to a state *Y* and their weights $\boldsymbol{\omega}_{X,Y}$

- *Aggregation characteristics*

  For any state *Y*, some combination function $\mathbf{c}_Y(..)$ defines the aggregation that is applied to the impacts $\boldsymbol{\omega}_{X,Y}X(t)$ on *Y* from its incoming connections from states *X*

- *Timing characteristics*

  Each state *Y* has a speed factor $\boldsymbol{\eta}_Y$ defining how fast it changes for given causal impact.

The following difference (or differential) equations that are used for simulation purposes and also for analysis of temporal-causal networks incorporate these network characteristics $\boldsymbol{\omega}_{X,Y}$, $\mathbf{c}_Y(..)$, $\boldsymbol{\eta}_Y$ in a standard numerical format:

$$Y(t+\Delta t) = Y(t) + \boldsymbol{\eta}_Y[\mathbf{c}_Y(\boldsymbol{\omega}_{X_1,Y}X_1(t), \ldots, \boldsymbol{\omega}_{X_k,Y}X_k(t)) - Y(t)]\Delta t \quad (4.1)$$

for any state *Y* and where X_1 to X_k are the states from which *Y* gets its incoming connections. Within the software environment described in (Treur 2020b, Ch. 9), a large number of around 40 useful basic combination functions are included in a combination function library.

The above concepts enable to design network models and their dynamics in a declarative manner, based on mathematically defined functions and relations. Realistic network models are usually adaptive: often not only their states but also some of their network characteristics change over time. By using a *self-modeling network* (also called a *reified* network), a similar network-oriented conceptualisation can also be applied to *adaptive* networks to obtain a declarative description using mathematically defined functions and relations for them as well; see (Treur 2018, 2020a, b). This works through the addition of new states to the network (called *self-model states*) which represent (adaptive) network characteristics. In the graphical 3D-format as shown in Sect. 4, such additional states are depicted at a next level (called *self-model level* or *reification level*), where the original network is at the *base level*. As an example, the weight $\boldsymbol{\omega}_{X,Y}$ of a connection from state *X* to state *Y* can be represented (at a next self-model level) by a self-model state named $\mathbf{W}_{X,Y}$ (objective representation actually used) or $\mathbf{RW}_{X,Y}$ (subjective representation for a person-related self-model). Similarly, all other network characteristics from $\boldsymbol{\omega}_{X,Y}$, $\mathbf{c}_Y(..)$, $\boldsymbol{\eta}_Y$ can be made adaptive by including self-model states for them. For example, an adaptive speed factor $\boldsymbol{\eta}_Y$ can be represented by a self-model state named $\mathbf{H}_Y$ and an adaptive excitability threshold parameter $\boldsymbol{\tau}_Y$ can be represented by a self-model state named $\mathbf{T}_Y$.

As the outcome of such a process of network reification is also a temporal-causal network model itself, as has been proven in (Treur 2020b, Ch 10), this self-modeling network construction can easily be applied iteratively to obtain multiple orders of self-models at multiple self-model levels. In the current chapter, a multi-level self-modeling network will be applied to obtain a second-order adaptive mental network model addressing metacognitive control of learning in a multiple mental models context.

4.4 A Mental Network Model for Metacognitive Control of Learning from Multiple Internal Mental Models

In this section, the adaptive mental network model for metacognitive control on learning using multiple mental models is introduced. This adaptive mental network model has processes at three levels:

- The base level network for the (multiple) internal mental models used
- The first-order self-model level for the learning of the internal mental models by adaptations of them
- The second-order self-model level for control by adaptation of the first-order network for the learning

These three levels of processes have been modeled by a second-order adaptive self-modeling network (Treur 2018, 2020a,b) briefly described in Sect. 3; the connectivity of this network model is depicted in Fig. 4.2. The states used are explained in Table 4.1. For the example mental models at the base level, on the left hand side in Fig. 4.2 an internal numerical mental model for an arithmetic task is included and on the right hand side a visual, geometrical mental model for it. The example task is to show (in the numerical representation) for certain given natural numbers *a*, *b* and *c* that

$$a^*(b+c) = a^*b + a^*c \tag{4.2}$$

Note that this is often applied in calculations, for example, when calculating 9*48 by splitting it as $9 * 40 + 9 * 8 = 360 + 72 = 432$, or by calculating 27*7 + 27*3 as 27*(7 + 3) = 270.

The detour via visualisation considers two rectangles with vertical dimension *a* and horizontal dimensions *b* and *c* and their areas that are together equal to the area of a rectangle with vertical dimension *a* and horizontal dimension *b* + *c*, as shown in Fig. 4.1.

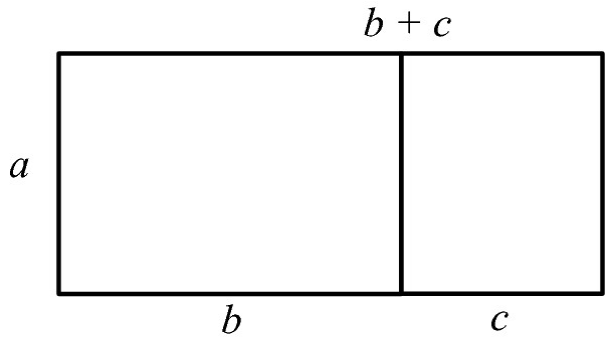


Fig. 4.1 Visualisation for the task expressed numerically by (2)

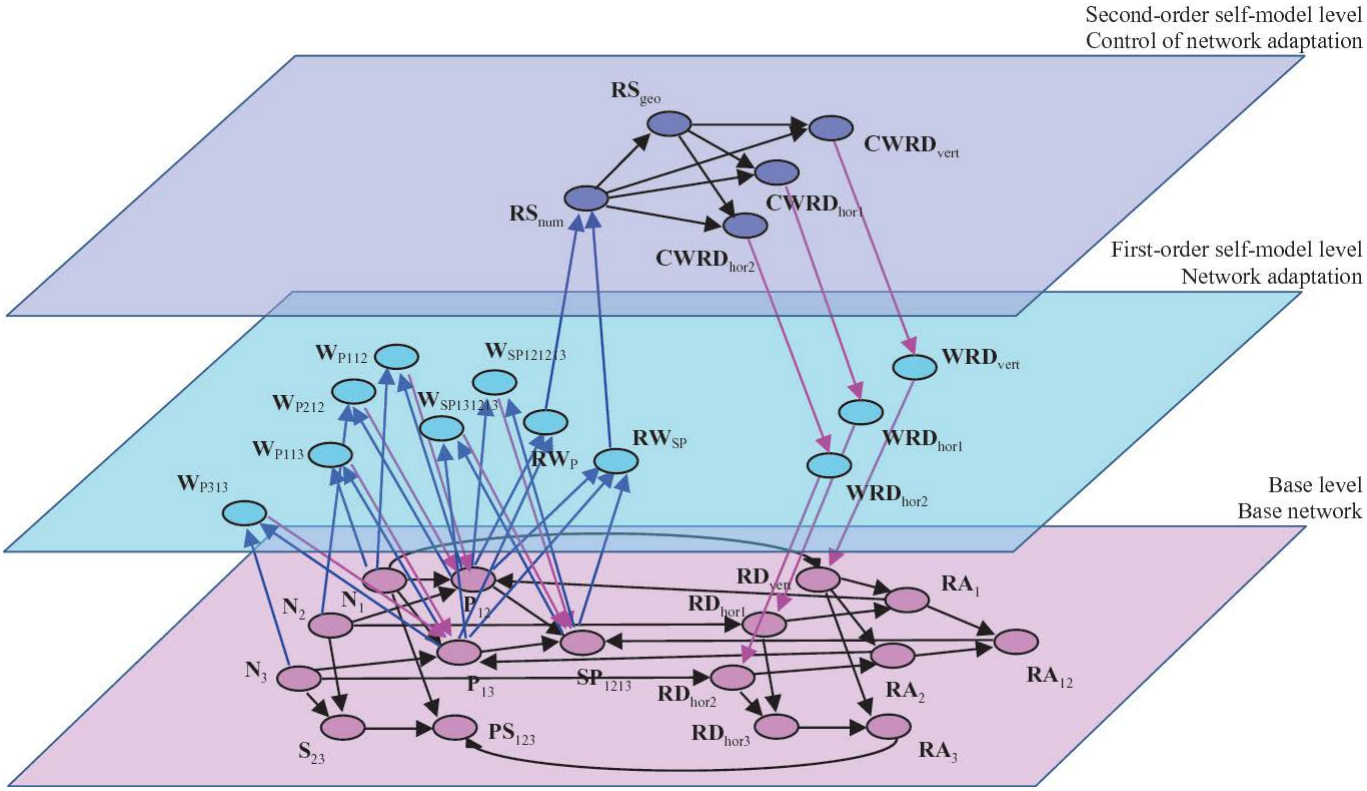


Fig. 4.2 Graphical representation of the connectivity of the second-order adaptive mental network model for metacognitive control of learning for multiple mental models

4.4.1 Network Characteristics: Connectivity and Timing

At the base level, for the numerical mental model, the base states $\mathbf{N}_1$, $\mathbf{N}_2$ and $\mathbf{N}_3$ represent the given numbers a, b, and c. Base states $\mathbf{P}_{12}$ and $\mathbf{P}_{13}$ represent the products $a*b$ and $a*c$, respectively, whereas state $\mathbf{S}_{12}$ represents the sum $b + c$. Finally, base state $\mathbf{SP}_{1213}$ represents the sum of $\mathbf{P}_{12}$ and $\mathbf{P}_{13}$ which is $a*b + a*c$, while base state $\mathbf{PS}_{123}$ represents the product of $\mathbf{N}_1$ and $\mathbf{S}_{23}$ which is $a*(b + c)$. For the geometric mental model, base states $\mathbf{RD}_{\text{vert}}$, $\mathbf{RD}_{\text{hor1}}$, $\mathbf{RD}_{\text{hor2}}$, and $\mathbf{RD}_{\text{hor3}}$ represent the vertical and horizontal dimensions of the rectangles in Fig. 4.1, respectively. Moreover, $\mathbf{RA}_1$, $\mathbf{RA}_2$ and $\mathbf{RA}_3$ represent the areas of the three

rectangles with horizontal dimension b, c, and $b + c$, respectively, and $\mathbf{RA}_{12}$ the area of the two smaller rectangles together.

At the first-order self-model level, the learning of the adaptive connections of the numerical mental model is modeled by the **W**-states and as input for the the self-model for the metacognitive monitoring the learnt relations as estimated by the learner are represented by the two (subjective) **RW**-states X_{23} and X_{24}. Moreover, the **WRD**-states X_{25} to X_{27} model the adaptive connections from the numerical mental model to the geometric mental model used to dynamically switch from one to the other; this is part of effectuating the metacognitive control.

At the second-order self-model level, the self-model for the status of the learning (for the own estimated learnt numerical and geometric skills) for the metacognitive monitoring is represented by the two **RS**-states X_{28} and X_{29} and the metacognitive control decisions (to switch to the geometric mental model) are modeled by the **CWRD**-states X_{30} to X_{32}, based on the impact from the self-model obtained by the metacognitive monitoring.

There are two types of connections: intra-level connections (in Fig. 4.2 depicted in black) and interlevel connections (depicted in blue for upward and in pink for downward). At the base level, within each of the two mental models, the connections define these mental models by their internal causal impacts. For example, the connections $\mathbf{N}_1 \rightarrow \mathbf{P}_{12}$ and $\mathbf{N}_2 \rightarrow \mathbf{P}_{12}$ define that within the numerical mental model the product of a and b represented by base state $\mathbf{P}_{12}$ depends on base states $\mathbf{N}_1$ and $\mathbf{N}_2$ representing these numbers.

In addition, at the base level a number of connections define how the two mental models relate to each other. For example, the connection $\mathbf{N}_1 \rightarrow \mathbf{RD}_{\text{vert}}$ from the numerical mental model to the geometric mental model defines that the vertical dimension of the rectangles within the geometric mental model depends on the number a represented by numerical state $\mathbf{N}_1$. Moreover, a connection back from the geometric to the numerical mental model such as $\mathbf{RA}_{12} \rightarrow \mathbf{SP}_{1213}$ defines the influence of the outcomes of the geometric process on the numerical process as a form of reinforcement to amplify the learning of the numerical mental model.

The upward connections to the first-order self-model **W**-states provide impact to the **W**-states so that they can adapt over time, which is modeled according to a qualitative Hebbian learning (Hebb 1949) principle specified by (4.6) below. For example, connections $\mathbf{N}_3 \rightarrow \mathbf{W}_{\text{P313}}$ and $\mathbf{P}_{13} \rightarrow \mathbf{W}_{\text{P313}}$ provide impact to $\mathbf{W}_{\text{P313}}$ so that $\mathbf{W}_{\text{P313}}$ can adapt over time. On the other hand, the downward connection from a **W**-state makes that the value of it is actually used

in the processing of the mental model. For example, the connection $\mathbf{W}_{P313} \rightarrow \mathbf{P}_{13}$ takes care for this for $\mathbf{W}_{P313}$ so that for the weight of the connection $\mathbf{N}_3 \rightarrow \mathbf{P}_{13}$ the value of $\mathbf{W}_{P313}$ is used. Furthermore, the upward connections to the first-order self-model **RW**-states make that a representation for the status of some connections of the numerical mental model is formed and maintained. This is a first step toward a self-model which is the basis of the metacognitive monitoring of the own cognitive processes.

At the second-order self-model level, based on impact from the **RW**-states at the first-order self-model level, the self-model is formed and maintained by the states $\mathbf{RS}_{num}$ and $\mathbf{RS}_{geo}$. Via their outgoing connections, the states $\mathbf{RS}_{num}$ and $\mathbf{RS}_{geo}$ of this self-model have their impact on the control decisions modeled by the **CWRD-**states. By their downward connections, the **CWRD-**states for control decisions determine the incoming connections to the corresponding **WRD**-states, so that the control decision is executed by realising that these **WRD**-states get values 1. In turn, once the **WRD**-state has a value 1, it makes that at the base level the corresponding connection from numerical mental model to geometric mental model is 1, which then leads to the geometric mental model states $\mathbf{RD}_{vert}$, $\mathbf{RD}_{hor1}$, and $\mathbf{RD}_{hor2}$ getting the appropriate values from states $\mathbf{N}_1$, $\mathbf{N}_2$, and $\mathbf{N}_3$ of the numerical mental model.

In Box 4.1 the complete role matrix specification of the connectivity and timing characteristics of the designed adaptive network model can be found. Here in each role matrix, each state has its row where it is listed which are the impacts on it from that role. Role matrix **mb** lists the other states (at the same or lower level) from which the state gets its incoming connections, whereas in role matrix **mcw** the connection weights are listed for these connections. Note that nonadaptive connection weights are indicated by a number (in a green shaded cell), but adaptive connection weights are indicated by a reference to the (self-model) state representing the adaptive value (in a peach-red shaded cell). For example, state X_5 (= $\mathbf{P}_{12}$) has incoming connections from X_1 (= $\mathbf{N}_1$), X_2 (= $\mathbf{N}_2$), and X_{13} (= $\mathbf{RA}_1$) with connection weights represented by X_{17} (= $\mathbf{W}_{P112}$) and X_{18} (= $\mathbf{W}_{P212}$) and 1, respectively. These two adaptive connection weights model the reinforced (by $\mathbf{RA}_1$) Hebbian learning. Also, the states $\mathbf{RD}_{vert}$, $\mathbf{RD}_{hor1}$, $\mathbf{RD}_{hor2}$ for the dimensions of the rectangles in the geometric mental model have adaptive connection weights. These adaptive connections are used to model the metacognitive control of the switch from numerical mental model to geometric mental model: if the control decision is made to switch, then these connection weights (represented by the **WRD**-states) quickly become 1 to transfer the numbers a, b and c to the geometric mental model. This rapid transition is specified in role matrix **ms** for the timing, where it is indicated that the speed

factors of the **WRD**-states X_{25} to X_{27} are adaptive and immediately change from 0 to 1 as soon as the **CWRD-**states X_{30} to X_{32} for metacognitive control at the second-order self-model level change to 1.

Box 4.1. Role matrices for the connectivity and timing characteristics of the network model

mb	base connectivity	1	2	3
X_1	$\mathbf{N}_1$	X_1		
X_2	$\mathbf{N}_2$	X_2		
X_3	$\mathbf{N}_3$	X_3		
X_4	$\mathbf{S}_{23}$	X_2	X_3	
X_5	$\mathbf{P}_{12}$	X_1	X_2	X_{13}
X_6	$\mathbf{P}_{13}$	X_1	X_3	X_{14}
X_7	$\mathbf{PS}_{123}$	X_1	X_4	X_{15}
X_8	$\mathbf{SP}_{1213}$	X_5	X_6	X_{16}
X_9	$\mathbf{RD}_{vert}$	X_1		
X_{10}	$\mathbf{RD}_{hor1}$	X_2		
X_{11}	$\mathbf{RD}_{hor2}$	X_3		
X_{12}	$\mathbf{RD}_{hor3}$	X_{10}	X_{11}	
X_{13}	$\mathbf{RA}_1$	X_9	X_{10}	
X_{14}	$\mathbf{RA}_2$	X_9	X_{11}	
X_{15}	$\mathbf{RA}_3$	X_9	X_{12}	
X_{16}	$\mathbf{RA}_{12}$	X_{13}	X_{14}	
X_{17}	$\mathbf{W}_{P112}$	X_1	X_5	X_{17}
X_{18}	$\mathbf{W}_{P212}$	X_2	X_5	X_{18}
X_{19}	$\mathbf{W}_{P113}$	X_1	X_6	X_{19}
X_{20}	$\mathbf{W}_{P313}$	X_3	X_6	X_{20}
X_{21}	$\mathbf{W}_{SP121213}$	X_5	X_8	X_{21}
X_{22}	$\mathbf{W}_{SP131213}$	X_6	X_8	X_{22}
X_{23}	$\mathbf{RW}_P$	X_5	X_6	
X_{24}	$\mathbf{RW}_{SP}$	X_5	X_6	X_8
X_{25}	$\mathbf{WRD}_{vert}$	X_{25}		
X_{26}	$\mathbf{WRD}_{hor1}$	X_{26}		
X_{27}	$\mathbf{WRD}_{hor2}$	X_{27}		
X_{28}	$\mathbf{RS}_{num}$	X_{23}	X_{24}	
X_{29}	$\mathbf{RS}_{geo}$	X_{28}		
X_{30}	$\mathbf{CWRD}_{vert}$	X_{25}	X_{29}	
X_{31}	$\mathbf{CWRD}_{hor1}$	X_{25}	X_{29}	
X_{32}	$\mathbf{CWRD}_{hor2}$	X_{25}	X_{29}	

mcw	connection weights	1	2	3
X_1	$\mathbf{N}_1$	1		
X_2	$\mathbf{N}_2$	1		
X_3	$\mathbf{N}_3$	1		
X_4	$\mathbf{S}_{23}$	1	1	
X_5	$\mathbf{P}_{12}$	X_{17}	X_{18}	1
X_6	$\mathbf{P}_{13}$	X_{19}	X_{20}	1
X_7	$\mathbf{PS}_{123}$	1	1	1
X_8	$\mathbf{SP}_{1213}$	X_{21}	X_{22}	1
X_9	$\mathbf{RD}_{vert}$	X_{25}		
X_{10}	$\mathbf{RD}_{hor1}$	X_{26}		
X_{11}	$\mathbf{RD}_{hor2}$	X_{27}		
X_{12}	$\mathbf{RD}_{hor3}$	1	1	
X_{13}	$\mathbf{RA}_1$	1	1	
X_{14}	$\mathbf{RA}_2$	1	1	
X_{15}	$\mathbf{RA}_3$	1	1	
X_{16}	$\mathbf{RA}_{12}$	1	1	
X_{17}	$\mathbf{W}_{P112}$	1	1	1
X_{18}	$\mathbf{W}_{P212}$	1	1	1
X_{19}	$\mathbf{W}_{P113}$	1	1	1
X_{20}	$\mathbf{W}_{P313}$	1	1	1
X_{21}	$\mathbf{W}_{SP121213}$	1	1	1
X_{22}	$\mathbf{W}_{SP131213}$	1	1	1
X_{23}	$\mathbf{RW}_P$	1	1	
X_{24}	$\mathbf{RW}_{SP}$	1	1	1
X_{25}	$\mathbf{WRD}_{vert}$	0		
X_{26}	$\mathbf{WRD}_{hor1}$	0		
X_{27}	$\mathbf{WRD}_{hor2}$	0		
X_{28}	$\mathbf{RS}_{num}$	1	1	
X_{29}	$\mathbf{RS}_{geo}$	1		
X_{30}	$\mathbf{CWRD}_{vert}$	-0.1	1	
X_{31}	$\mathbf{CWRD}_{hor1}$	-0.1	1	
X_{32}	$\mathbf{CWRD}_{hor2}$	-0.1	1	

ms	speed factors	1
X_1	$\mathbf{N}_1$	0
X_2	$\mathbf{N}_2$	0
X_3	$\mathbf{N}_3$	0
X_4	$\mathbf{S}_1$	0.5
X_5	$\mathbf{P}_{12}$	0.5
X_6	$\mathbf{P}_{13}$	0.5
X_7	$\mathbf{PS}_{123}$	0.5
X_8	$\mathbf{SP}_{1213}$	0.5
X_9	$\mathbf{RD}_{vert}$	0.5
X_{10}	$\mathbf{RD}_{hor1}$	0.5
X_{11}	$\mathbf{RD}_{hor2}$	0.5
X_{12}	$\mathbf{RD}_{hor3}$	0.5
X_{13}	$\mathbf{RA}_1$	0.5
X_{14}	$\mathbf{RA}_2$	0.5
X_{15}	$\mathbf{RA}_3$	0.15
X_{16}	$\mathbf{RA}_{12}$	0.05
X_{17}	$\mathbf{W}_{P112}$	0.02
X_{18}	$\mathbf{W}_{P212}$	0.02
X_{19}	$\mathbf{W}_{P113}$	0.02
X_{20}	$\mathbf{W}_{P313}$	0.02
X_{21}	$\mathbf{W}_{SP121213}$	0.02
X_{22}	$\mathbf{W}_{SP131213}$	0.02
X_{23}	$\mathbf{RW}_P$	0.1
X_{24}	$\mathbf{RW}_{SP}$	0.1
X_{25}	$\mathbf{WRD}_{vert}$	X_{30}
X_{26}	$\mathbf{WRD}_{hor1}$	X_{31}
X_{27}	$\mathbf{WRD}_{hor2}$	X_{32}
X_{28}	$\mathbf{RS}_{num}$	0.1
X_{29}	$\mathbf{RS}_{geo}$	0.5
X_{30}	$\mathbf{CWRD}_{vert}$	0.5
X_{31}	$\mathbf{CWRD}_{hor1}$	0.5
X_{32}	$\mathbf{CWRD}_{hor2}$	0.5

4.4.2 Network Characteristics: Aggregation

The network characteristics for aggregation are defined by the selection of combination functions from the library and values for their parameters. First the six combination functions used for the model are specified by

$$\mathbf{mcf} = [1, 2, 39, 22, 23, 4]$$

$$= [\mathbf{eucl}, \mathbf{alogistic}, \mathbf{hebbqual}, \mathbf{complement-id}, \mathbf{product}, \mathbf{max-composition}]$$

Here the numbers are the numbers of the listed functions in the library. Next, it is specified which state uses which combination function. This can be seen in role matrix **mcfw** in Box 4.2.

Box 4.2. Role matrices for the aggregation characteristics: combination functions and their parameters

mcfw combination weights		1 eucl	2 alog-istic	3 hebb-qual	4 comp-id	5 pro-duct	6 max-comp
X_1	N_1	1					
X_2	N_2	1					
X_3	N_3	1					
X_4	S_{23}	1					
X_5	P_{12}						1
X_6	P_{13}						1
X_7	PS_{123}						1
X_8	SP_{1213}						1
X_9	RD_{vert}	1					
X_{10}	RD_{hor1}	1					
X_{11}	RD_{hor2}	1					
X_{12}	RD_{hor3}	1					
X_{13}	RA_1					1	
X_{14}	RA_2					1	
X_{15}	RA_3					1	
X_{16}	RA_{12}	1					
X_{17}	WP_{112}			1			
X_{18}	WP_{212}			1			
X_{19}	WP_{113}			1			
X_{20}	WP_{313}			1			
X_{21}	WSP_{121213}			1			
X_{22}	WSP_{131213}			1			
X_{23}	RW_P		1				
X_{24}	RW_{SP}		1				
X_{25}	WRD_{vert}				1		
X_{26}	WRD_{hor1}				1		
X_{27}	WRD_{hor2}				1		
X_{28}	RS_{num}		1				
X_{29}	RS_{geo}				1		
X_{30}	$CWRD_{vert}$		1				
X_{31}	$CWRD_{hor1}$		1				
X_{32}	$CWRD_{hor2}$		1				

mcfp combination function parameters		1 eucl		2 alog-istic		3 hebb-qual		4 comp-id		5 pro-duct		6 max-comp	
		1 n	2 λ	1 σ	2 τ	1 μ	2	1	2	1	2	1 γ_1	2 γ_2
X_1	N_1	1	1										
X_2	N_2	1	1										
X_3	N_3	1	1										
X_4	S_{23}	1	1										
X_5	P_{12}											23	1
X_6	P_{13}											23	1
X_7	PS_{123}											23	1
X_8	SP_{1213}											1	1
X_9	RD_{vert}	1	1										
X_{10}	RD_{hor1}	1	1										
X_{11}	RD_{hor2}	1	1										
X_{12}	RD_{hor3}	1	1										
X_{13}	RA_1												
X_{14}	RA_2												
X_{15}	RA_3												
X_{16}	RA_{12}	1	1										
X_{17}	WP_{112}					1							
X_{18}	WP_{212}					1							
X_{19}	WP_{113}					1							
X_{20}	WP_{313}					1							
X_{21}	WSP_{121213}					1							
X_{22}	WSP_{131213}					1							
X_{23}	RW_P			8	1.5								
X_{24}	RW_{SP}			8	2								
X_{25}	WRD_{vert}												
X_{26}	WRD_{hor1}												
X_{27}	WRD_{hor2}												
X_{28}	RS_{num}			8	1.5								
X_{29}	RS_{geo}												
X_{30}	$CWRD_{vert}$			18	0.2								
X_{31}	$CWRD_{hor1}$			18	0.2								
X_{32}	$CWRD_{hor2}$			18	0.2								

The combination functions from the library used in the introduced network model are defined as follows:

- The *Euclidean combination function* $\mathbf{eucl_{n,\lambda}}(V_1, ..., V_k)$ is defined by

$$\mathbf{eucl}_{n,\lambda}(V_1, \ldots, V_k) = \sqrt[n]{V_1^n + \ldots + V_k^n} \tag{4.3}$$

where $\boldsymbol{n}$ is the order and $\boldsymbol{\lambda}$ a scaling factor and $V_1, ..., V_k$ are the impacts from the states from which the considered state Y gets incoming connections. Note that if both parameters have value 1, then this is just the sum function and when there is only one incoming connection the identity function. This is always the case in the current model, as can be seen in role matrix **mcfp**.

- The *product combination function* $\mathbf{product}(V_1, V_2)$ is defined by

$$\mathbf{product}(V_1, V_2) = V_1 V_2 \tag{4.4}$$

- The *advanced logistic sum combination function* $\mathbf{alogistic_{\sigma,\tau}}(V_1, ..., V_k)$ is defined by:

$$\mathbf{alogistic}_{\sigma,\tau}(V_1, \ldots, V_k) = \left[\frac{1}{1 + \mathrm{e}^{-\sigma(V_1 + \ldots + V_k - \tau)}} - \frac{1}{1 + \mathrm{e}^{\sigma,\tau)}}\right]\left(1 + \mathrm{e}^{-\sigma,\tau}\right) \tag{4.5}$$

where $\boldsymbol{\sigma}$ is a steepness parameter and $\boldsymbol{\tau}$ a threshold parameter and $V_1, ..., V_k$ are the impacts from the states from which the considered state Y gets incoming connections

- The *qualitative Hebbian learning combination function* $\mathbf{hebbqual_{\mu}}(V_1, V_2, W)$ is defined by

$$\mathbf{hebbqual}(V_1, V_2, W) = V_1^* V_2^* (1 - W) + \mu W \tag{4.6}$$

where $\boldsymbol{\mu}$ is a persistence parameter, W represents the weight of their connection, and V_i^* is 1 if $V_i > 0.1$ and else 0 (here V_1, V_2 are the activation levels of the connected states).

- The *complemental identity combination function* $\mathbf{complement\text{-}id}(V)$ is defined by

$$\mathbf{complement - id}(V) = 1 - V \tag{4.7}$$

where V is the incoming impact from a connected state

- The *max-composing combination function* $\mathbf{max\text{-}composition_{m,n}}(V_1, V_2, V_3)$ is defined by

$$\mathbf{max - composition}_{m,n}(V_1, V_2, V_3) = \max(\mathrm{bcf}(\boldsymbol{m}, [1,1], [V_1, V_2]), \mathrm{bcf}(\boldsymbol{n}, [1,1], [V_3])) \tag{4.8}$$

where bcf(i,p,v) is the i^{th} basic combination function from the library. This function composes two other combination functions from the library by using

the max-function. It is actually defined as a special case using a more general function available in the combination function library that enables to create any function composition of any combination functions from the library: the function

$$\mathbf{composedbcfs}\,(h, p, nrs, ps, vs, ks)$$

which is defined as a function

$$\mathrm{bcf}\,(h, p,\ \mathrm{bcfvalues}\,(nrs, ps, vs, ks))$$

where *m* = length(*nrs*) is the number of functions composed with function number *h*, *p* is a list of parameter values of the composing function number *h*, *nrs* a list for the numbers of the composed functions, *ps* for their parameters, *vs* for their values and *ks* the numbers of their arguments, and (assuming two parameters per function)

$$\begin{aligned}\mathrm{bcfvalues}\,(nrs, ps, vs, ks) = [&\mathrm{bcf}\,(nrs\,(1)\,, [ps\,(1)\,, ps\,(2)]\,, [vs\,(1)\,, \ldots, v\,(ks\,(1))]\,, \ldots,\\ &\mathrm{bcf}\,(nrs\,(m)\,, [ps\,(2m-1)\,, ps\,(2m)]\,, \left[vs\left(1+\Sigma_{i=1}^{m-1} ks\,(i)\right), \ldots,\right.\\ &\left. vs\left(\Sigma_{i=1}^{m} ks\,(i)\right)\right])]\end{aligned}$$

The combination function **eucl$_{n,\lambda}$(...)** for ***n*** and **λ** both 1, is used to model addition, **product(**V_1, V_2**)** to model multiplication, **hebbqual$_{\mu}$(**V_1, V_2, W**)** to model learning of arithmetic operations, and **alogistic$_{\sigma,\tau}$(**V_1, ..., V_k**)** and **complement-id(**V**)** to model internal metacognitive monitoring and control states for the learning. The combination function **max-composition$_{m,n}$(**V_1, V_2, V_3**)** is used to reinforce the learning in the numerical mental model through the outcomes from the geometric mental model.

4.5 Example Simulation Scenarios

In this section, simulations of two example scenarios will be discussed to illustrate the introduced second-order adaptive network model. Both scenarios address the example task discussed in Sect. 3 (see also Fig. 4.1) for *a* = 2, *b* = 3, *c* = 2, which are used as constant values for base states $\mathbf{N}_1$, $\mathbf{N}_2$, and $\mathbf{N}_3$, respectively. The first scenario shows how someone who has good arithmetic skills addresses the task, without involving any switch to the geometric mental model; see Fig. 4.3. As can be seen, as one of the first, state $\mathbf{S}_{23}$ comes up which determines the sum of $\mathbf{N}_2$ representing *b* and $\mathbf{N}_3$ representing *c*, which correctly ends up in value 5 (the blue line). At about the same time state $\mathbf{P}_{12}$ (the red line) for the product of $\mathbf{N}_1$ and $\mathbf{N}_2$ representing *a* and *b* comes up, correctly ending up at 6.

Similarly, $\mathbf{P}_{13}$ (the blue-green line) for the product of $\mathbf{N}_1$ (for a) and $\mathbf{N}_2$ (for c) correctly reaches 4.

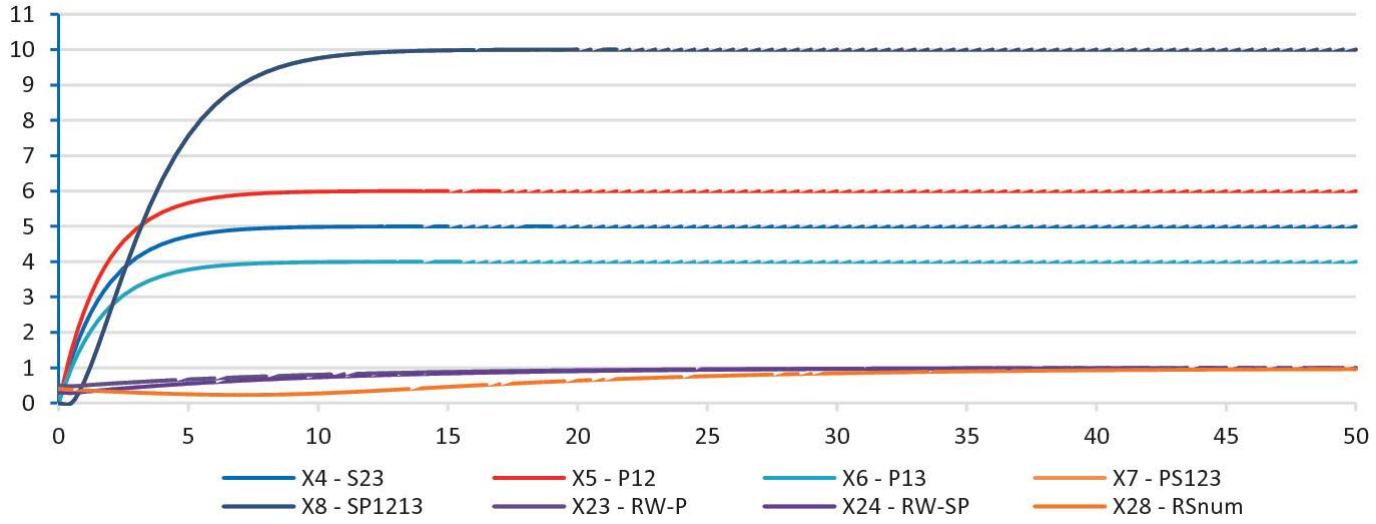


Fig. 4.3 Using the arithmetic mental model and formation of the self-model for metacognitive monitoring

Next, $\mathbf{PS}_{123}$ of $\mathbf{N}_1$ and $\mathbf{S}_{23}$ representing the product of a and $b + c$ is determined, which correctly ends up in 10 (the light dark green line). The determines the left hand side of the Eq. (4.2). At the same time, the right hand side of (2) is addressed. Therefore, $\mathbf{SP}_{1213}$ (again the dark green line) for the sum of $\mathbf{P}_{12}$ and $\mathbf{P}_{13}$ comes up and correctly reaches 10. This shows that the right hand side of (2) is indeed equal to the left hand side of (2), what solves the task. In the meantime it can be seen in Fig. 4.3 that the self-model about the numerical mental model is formed: the two lines for the two **RW**-states all end up at 1, and also based on them the third (orange) line for $\mathbf{RS}_{num}$, which as a form of metacognitive monitoring tells the learner that the arithmetic skills are OK. Therefore, in this case no control decision to switch to the geometric mental model is made, and also no further learning is needed.

The second scenario is the more interesting one (see Fig. 4.4). Here the learner has still good arithmetic skills (connection weights 1) to address the left hand side of (2), but not for the right hand side (connection weights are only 0.1). Therefore the light brown and purple lines in the upper graph in Fig. 4.4 are the same as in Fig. 4.3, but not the lines for $\mathbf{P}_{12}$, $\mathbf{P}_{13}$, and $\mathbf{SP}_{1213}$ needed for the right hand side of (2). Because that side gets stuck, and the self-model used for monitoring has low values showing a lack of arithmetic skills, the control decision is made to switch to the geometric mental model: all three **CWRD**-states come up soon and reach 1 shortly after time 5 (the purple line in the lower graph of Fig. 4.4). As a consequence, to execute this control decision, the **WRD**-states become 1 around time 5 (the red line in the lower graph of Fig. 4.4).

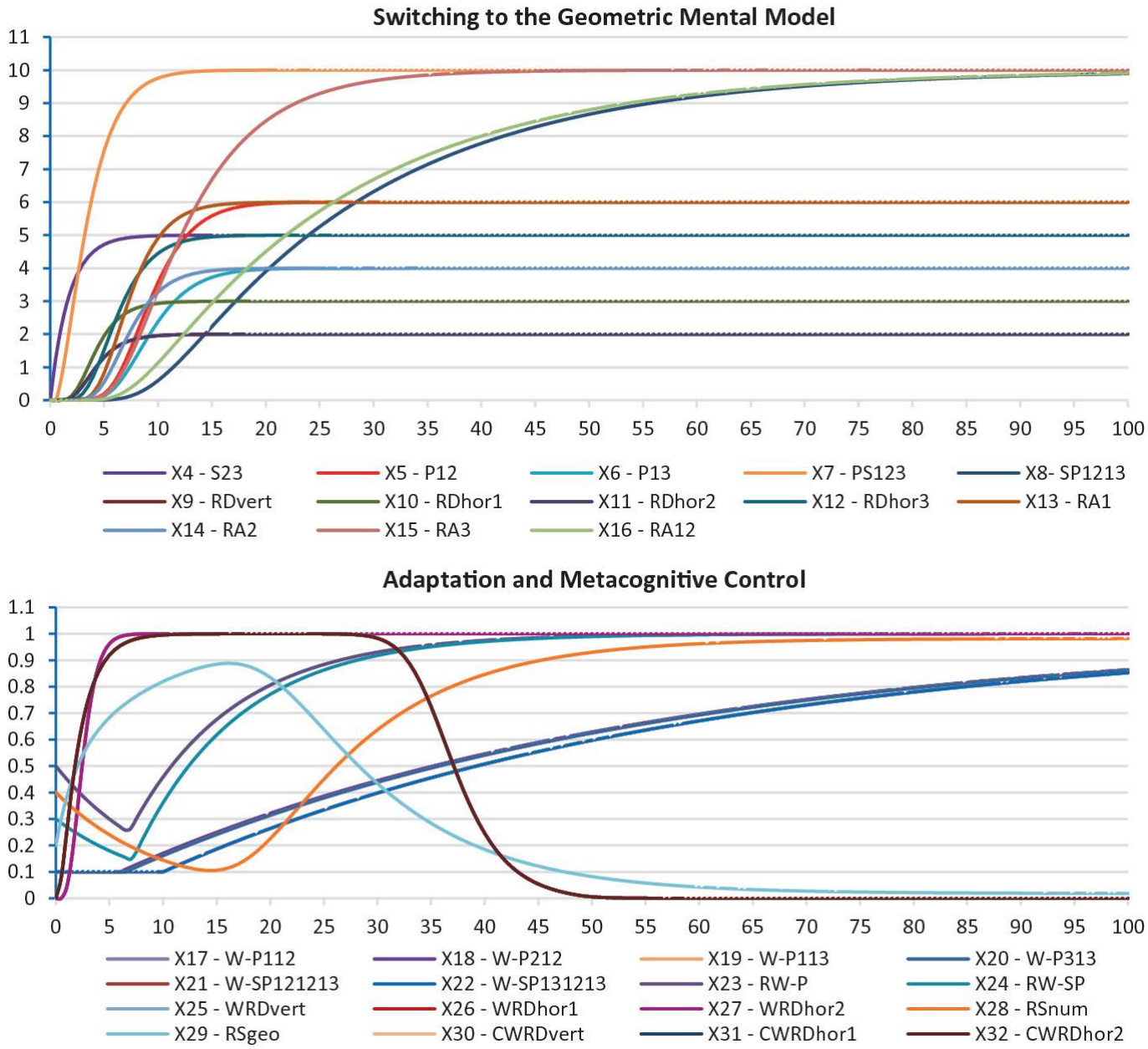


Fig. 4.4 Switching to the geometric mental model and adaptation and metacognitive control of it

Because of that the **RD**-states representing the dimensions of the rectangles get their values 2, 3, and 5. Based on these, the **RA**-states for the areas of the rectangles are determined and get their values 4, 6 and 10. As these **RA**-states provide a reinforcing impact on the states $\mathbf{P}_{12}$, $\mathbf{P}_{13}$, and $\mathbf{SP}_{1213}$ in the numerical mental model, it can be seen that with a small delay the latter states follow the **RA**-states to also reach values 4, 6 and 10 (the red, light blue, and dark green line in Fig. 4.4, upper graph). In the lower graph of Fig. 4.4 it can be seen what happens further concerning the adaptation levels. The lines starting at 0.1 are the **W**-states, and it is shown that after time 6 they start to increase to finally reach values close to 1. This is the reinforced Hebbian learning process for the numerical mental model: reinforced by the impact from the geometric mental

model. Also the two **RW**-states and state $\mathbf{RS}_{num}$ for the self-model for the numerical mental model, starting at 0.3, 0.5 and 0.4, increase after time 6. Note that the $\mathbf{RS}_{geo}$ (light green line with peak near 0.9) also increases thereby supporting the decision to switch to the geometric mental model, but later on (after time 15) goes down just like the **CWRD**-states for the control themselves do (after time 25), as after learning the full arithmetic mental model, by the monitoring via the self-model the learner feels that there is no reason anymore to consider switching to the geometric mental model.

4.6 Discussion

Learning processes can be described by adaptive mental (or neural) network models. If metacognition is used to regulate learning (Pintrich 2000), the adaptation of the mental network becomes itself adaptive as well, so then it involves second-order adaptation. In this chapter, a second-order adaptive mental network model was introduced for metacognitive regulation of learning processes using multiple internal mental models. Part of the material was adopted from (Treur 2021).

The focus was on the role of multiple mental models (Gentner and Stevens 1983; Greca and Moreira 2000; Skemp 1971; Seel 2006), in particular, the case of visualisation to support learning of numerical or symbolic skills (Bruner 1966, 1977; Bidwell 1972; Day and Hurell 2015; Du Plooy 2016; Freudenthal 1973,1986; Koedinger and Terao 2002; Larbi and Mavis 2016; Lovitt et al. 1984; Renkema 2019; Roberts 1989). The second-order adaptive network model was illustrated for the role of visualisation to support learning multiplication at the primary school.

It was shown how a second-order self-modeling network model provides adequate means to model the different aspects that make the addressed topic complex: the network has a self-model about its own structure, it models mental models and their adaptation for learning, and it models dynamic metacognitive control of this adaptation. The model was applied to simulate some example scenarios that illustrate what the model does. In further work other scenarios can be addressed as well.

References

Bidwell, J.K.: A physical model for factoring quadratic polynomials. Math. Teach. **65**(3), 201–205 (1972) [Crossref]

Bruner, J.S.: Towards a theory of instruction. Harvard University, Cambridge, Mass (1966)

Bruner, J.S.: The Process of Education. Harvard University Press, London (1977)

Darling-Hammond, L., Austin, K., Cheung, M., Martin, D.: Thinking about thinking: Metacognition. (2008)

Day, L., Hurrell, D.: An explanation for the use of arrays to promote the understanding of mental strategies for multiplication. Aust. Prim. Math. Classr. **20**(1), 20–23 (2015)

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**.** https://repository.up.ac.za/handle/2263/51258 (2016)

Flavell, J.H.: Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. Am. Psychol. **34**(10), 906–911 (1979)
[Crossref]

Freudental, H.: Mathematics as an Educational Task. Reidel, Dordrecht (1973)

Freudental, H.: Didactical Phenomenology of Mathematical Structures. Reidel, Dordrecht (1986)

Gentner, D., Stevens, A.L.: Mental models. Erlbaum, Hillsdale NJ (1983)

Greca, I.M., Moreira, M.A.: Mental models, conceptual models, and modelling. Int. J. Sci. Educ. **22**(1), 1–11 (2000)
[Crossref]

Hebb, D.O.: The organization of behavior: a neuropsychological theory. Wiley (1949)

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'02, pp. 542–547. Lawrence Erlbaum Associates, Mahwah, NJ (2002)

Koriat, A.: Metacognition and consciousness. In: Zelavo, P.D., Moscovitch, M., Thompson, E. (eds.). Cambridge Handbook of Consciousness. New York, Cambridge University Press (2007)

Larbi, E., Mavis, O.: The Use of manipulatives in mathematics education. J. Educ. Pract. **7**(36), 53–61 (2016)

Lovitt, C., Marriot, C., Swan, K.: Lessons in algebra using algebra blocks. Laverton, Victoria: EDU-DOMES (1984)

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

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).

Renkema, W.: Effectiveness of Dynamic Visualisation in Video-based Animated Algebra Instruction. Thesis, Utrecht University, M.Sc (2019)

Rivera, F.D.: Visualization and Progressive Schematization: Framing the Issues. In: Toward a Visually-Oriented School Mathematics Curriculum. Mathematics Education Library, vol 49, pp. 21–58. Springer, Dordrecht. https://doi-org.vu-nl.idm.oclc.org/https://doi.org/10.1007/978-94-007-0014-7_2 (2011).

Rivera, F.D.: Toward a visually-oriented school mathematics curriculum: research, theory, practice, and issues. Math. Educ. Library **49**. https://doi.org/10.1007/978-94-007-0014-7_2 (2011)

Roberts, B.S.: The effects of algebra blocks on student achievement in algebra. B.Sc.Thesis, Edith Cowan University. https://ro.ecu.edu.au/theses_hons/173 (1989)

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

Shannon, S.V.: Using metacognitive strategies and learning styles to create self-directed learners. Inst. Learning Styles J. **1**, 14–28 (2008)

Skemp, R.R.: The Psychology of Learning Mathematics. Penguin Books, Harmondsworth (1971)

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive. Springer Publishers, Affective and Social Interactions (2016)
[Crossref]

Treur, J.: Multilevel network reification: representing higher order adaptivity in a network. In: Aiello, L., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L. (eds), Proc. of the 7th Int. Conf. on Complex Networks and their Applications, ComplexNetworks'18, vol. 1. Studies in Computational Intelligence, vol. 812, pp. 635–651, Springer Nature (2018)

Treur, J.: Modeling higher-order adaptivity of a network by multilevel network reification. Network Science **8**, S110–S144 (2020a)
[Crossref]

Treur, J.: Network-oriented modeling for adaptive networks: Designing Higher-order Adaptive Biological, Mental and Social Network Models. Springer Nature Publishing, Cham, Switzerland (2020b)
[Crossref]

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

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_5

5. Disturbed by Flashbacks: A Controlled Adaptive Network Model Addressing Mental Models for Flashbacks from PTSD

Laila van Ments[1] ✉ and Jan Treur[2] ✉
(1) AutoLeadStar, Jerusalem, Israel
(2) Social AI Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

✉ **Laila van Ments (Corresponding author)**
Email: laila@autoleadstar.com

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
In this chapter, a second-order adaptive network model is introduced for a number of phenomena that occur in the context of PTSD. First of all the model covers simulation of the formation of a mental model of a traumatic course of events and its emotional responses that make replay of flashback movies happen. Secondly, it addresses learning processes of how a stimulus can become a trigger to activate this acquired mental model. Furthermore, the influence of therapy on the ability of an individual to learn to control the emotional responses to the traumatic mental model was modeled. Finally, a form of second-order adaptation was covered to unblock and activate this learning ability.

Keywords PTSD – Higher-order adaptive – Mental model – Flashback movie

5.1 Introduction

A Post Traumatic Stress Disorder (PTSD) is usually developed after experiencing an event (often consisting of a sequence of steps) that triggers strong negative emotions like fear; e.g., (Duvarci and Pare 2014; Parsons and Ressler 2013). One

of the symptoms is a recurring re-experiencing of the event sequence that led to the trauma and that are played again and again in the mind as a kind of flashback movie and thereby trigger the strong negative emotions again. The occurrence of such flashbacks can be described as a mental model that was learned during the traumatic event sequence and that is replayed by internal (mental) simulation.

In the literature such as (Admon et al. 2013; Akiki et al. 2017; Holmes et al. 2018; Zandvakili et al. 2020) strong evidence can be found for relations to amygdala, dorsal anterior cingulated cortex, ventromedial prefrontal cortex and hippocampus. One of the reported issues here is a reduction of the connections to regions of the prefrontal cortex, which makes it difficult to apply emotion regulation. The role of the amygdala in activating fear and of the relation between amygdala and the pre-frontal cortex areas in suppressing fear was found to be crucial; e.g., (Admon et al. 2013; Panksepp and Biven 2012). If the emotion regulation strategy based on suppression is strengthened, this leads to a decrease in physiological and experiential effects of negative emotions; e.g., (Fitzgerald et al. 2018; Ochsner and Gross 2014; Webb et al. 2012).

Multiple forms of adaptivity play a crucial role in both the development of PTSD and therapies to recover from it. During the development, an important role is played by the learning of a form of mental model of the event sequence leading to the trauma. This is a form of observational learning; e.g., (Benbassat 2014; Van Gog et al. 2009). It is this learnt mental model that is the basis of the flashback symptoms. Moreover, during development also learning takes place to connect different stimuli (by themselves irrelevant but just co-occurring with the traumatic events) to the traumatic stimuli which makes them triggers for the flashbacks; this is a form of sensory preconditioning; e.g., (Brogden 1947; Hall 1996). To recover from PTSD, another form of learning is required: learning to strengthen the connections to the relevant prefrontal cortex areas to improve emotion regulation; e.g., (Ochsner and Gross 2014; Webb et al. 2012). However, this learning capability is impaired by the stress itself, which prevents the learning from taking place in a natural manner. This effect is called metaplasticity; e.g., (Garcia 2002). Metaplasticity (Abraham and Bear 1996) is a form of second-order adaptation, as it exerts a form of control over adaptation. In contrast, the other forms of adaptation mentioned above are called first-order adaptation.

The focus in the current chapter is to introduce a computational network model addressing all these forms of adaptivity pointed out above. This leads to a second-order adaptive network model in which during development of PTSD a mental model for the flashbacks is learnt and also an association of a trigger to the traumatic events (both first-order adaptation). As an additional effect of the development phase, a negative effect of metaplasticity occurs that impairs the plasticity of the emotion regulation (second-order adaptation). For recovery, a

therapy is applied to resolve the impairment of the plasticity of the emotion regulation which is a positive effect of metaplasticity (second-order adaptation). After this, the learning to strengthen the emotion regulation takes place which then leads to recovery (first-order adaptation).

In Sect. 5.2 some background knowledge is discussed for the different types of adaptation. Section 5.3 introduces the second-order adaptive network model to address these forms of adaptation. In Sect. 5.4 some example simulations for this network model are discussed. Finally, Sect. 5.5 is a discussion.

5.2 Background Knowledge on Adaptation Principles Used

As discussed above, different forms of adaptation play a role in development of and recovery from traumas. The more specific adaptation principles for these forms of adaptation are discussed in this section.

5.2.1 First-Order Adaptation Principle: Hebbian Learning

In neuroscientific literature such as (Chandra and Barkai 2018), two types of first-order adaptation principles are discussed: synaptic and non-synaptic. An example of the latter type is intrinsic excitability adaptation, which will not be used here. Hebbian learning is a well-known first-order adaptation principle of the first type; it addresses adaptive connectivity (Hebb 1949). It can be explained by:

> 'When an axon of cell A is near enough to excite B and repeatedly or persistently (5.1)
> takes part in firing it, some growth process or metabolic change takes place in one
> or both cells such that A's efficiency, as one of the cells firing B, is increased.'
> (Hebb 1949), p. 62

This is sometimes simplified (neglecting the phrase 'one of the cells firing B') to:

> 'What fires together, wires together' (Keysers and Gazzola 2014; Shatz 1992) (5.2)

This first-order adaptation principle will be used to model adaptation for the following.

- Development of the trauma:
  - Learning of a connection of a trigger stimulus to the traumatic event sequence based on sensory preconditioning (Brogden 1947; Hall 1996).
  - Learning the connections in the mental model of the traumatic event sequence based on observational learning, also using sensory preconditioning (Benbassat 2014; Van Gog et al. 2009).
- Recovery from the trauma:
  - Strengthening emotion regulation for recovery by learning the connections to the prefrontal cortex areas (Ochsner and Gross 2014; Webb et al. 2012).

5.2.2 Second-Order Adaptation Principle: Stress Reduces Adaptation Speed

In (Garcia 2002) the focus is on the role of stress in reducing or blocking plasticity. Many mental and physical disorders are stress-related, and are hard to overcome due to poor or even blocked plasticity that comes with the stress. Garcia (2002) describes the negative role of stress-related metaplasticity for this, which often leans to a situation that a patient is locked in his or her disorder by that negative pattern. However, he also shows that by some form of therapy this negative cycle might be broken:

> 'At the cellular level, evidence has emerged indicating neuronal atrophy and cell loss in response to stress and in depression. At the molecular level, it has been suggested that these cellular deficiencies, mostly detected in the hippocampus, result from a decrease in the expression of brain-derived neurotrophic factor (BDNF) associated with elevation of glucocorticoids.' (Garcia 2002), p. 629
>
> '…modifications in the threshold for synaptic plasticity that enhances cognitive function is referred here to as 'positive' metaplasticity. In contrast, changes in the threshold for synaptic plasticity that yield impairment of cognitive functions, for example (..) in response to stress (..), is referred to as 'negative' metaplasticity.' (Garcia 2002), pp. 630–631
>
> 'In summary, depressive-like behavior in animals and human depression are associated with high plasma levels of glucocorticoids that produce 'negative' metaplasticity in limbic structures (…). This stress-related metaplasticity impairs performance on certain hippocampal-dependent tasks. Antidepressant treatments act by increasing expression of BDNF in the hippocampus. This antidepressant effect can trigger, in turn, the suppression of stress-related metaplasticity in hippocampal-

hypothalamic pathways thus restoring physiological levels of glucocorticoids.' (Garcia 2002), p. 634

This second-order adaptation principle will be used to model adaptation for the following.

- Development of the trauma:
  - Reducing the adaptation speed for the learning of the emotion regulation connections to the prefrontal cortex areas due to the high stress levels (Garcia 2002)
- Recovery from the trauma:
  - Increasing the adaptation speed for the learning of the emotion regulation connections to the prefrontal cortex areas due to a therapy that (temporarily) reduces the stress levels (Garcia 2002)

In Sect. 5.3 it will be discussed how these have been modeled by using a so-called self-modeling network model.

5.3 The Second-Order Adaptive Network Model

In this section, a detailed overview is presented of the designed second-order adaptive network model for modeling the learning of PTSD trauma and the influence of therapy on recovery. For the modeling, we use the Network-Oriented Modeling approach introduced in (Treur 2016) and further developed to cover higher-order adaptive networks in (Treur 2020a; b), where also the supporting dedicated software environment is presented.

5.3.1 The General Format

This approach can be broken down in the following steps:

- Translating the domain into a conceptual causal network model in terms of network characteristics
- Transcribing the conceptual causal network model into a standard table format called *role matrix format*. These role matrices break down the network characteristics for all the different types of causal influences on a state in the model
- The network characteristics are grouped into the following types:

1. **Connectivity characteristics**
   What *states* X, Y and *connections* $X \rightarrow Y$ are there in the model and what are the *weights* $\boldsymbol{\omega}_{X,Y}$ of the connections? These are specified in role matrix **mb**

(for the states and their connections) and **mcw** (for the connection weights $\boldsymbol{\omega}_{X,Y}$).

2. **Aggregation characteristics**
   How are different impacts from other states on a state *Y* aggregated by a *combination function* $\mathbf{c}_Y\mathbf{(..)}$ and what are the values of the *parameters* for these combination functions? The combination functions are chosen from a library by assigning weights $\boldsymbol{\gamma}_{i,Y}$ to them and values for the parameters $\boldsymbol{\pi}_{i,j,Y}$ are set. These characteristics are specified in role matrix **mcfw** (for combination function weights $\boldsymbol{\gamma}_{i,Y}$) and **mcfp** (for the combination function parameters $\boldsymbol{\pi}_{i,j,Y}$).
3. **Timing characteristics**
   How fast do the states *Y* change upon the received impact, due to their *speed factor* $\boldsymbol{\eta}_Y$? These speed factors $\boldsymbol{\eta}_Y$ are specified in role matrix **ms**.

- Providing the above network characteristics as tables in role matrix format as input for the available dedicated software environment. Based on these received tables, the software environment runs simulations.

5.3.2 Translating the Domain Knowledge into a Conceptual Causal Model

Based on a domain study, the first step towards building a computational model is translating the processes and brain mechanisms discussed in the literature into a conceptual causal network model. To accommodate for the forms of adaptation of different orders order for the model, the conceptual model uses so-called *self-modeling networks* that include self-models, in this case leading to three levels (see Fig. 5.1):

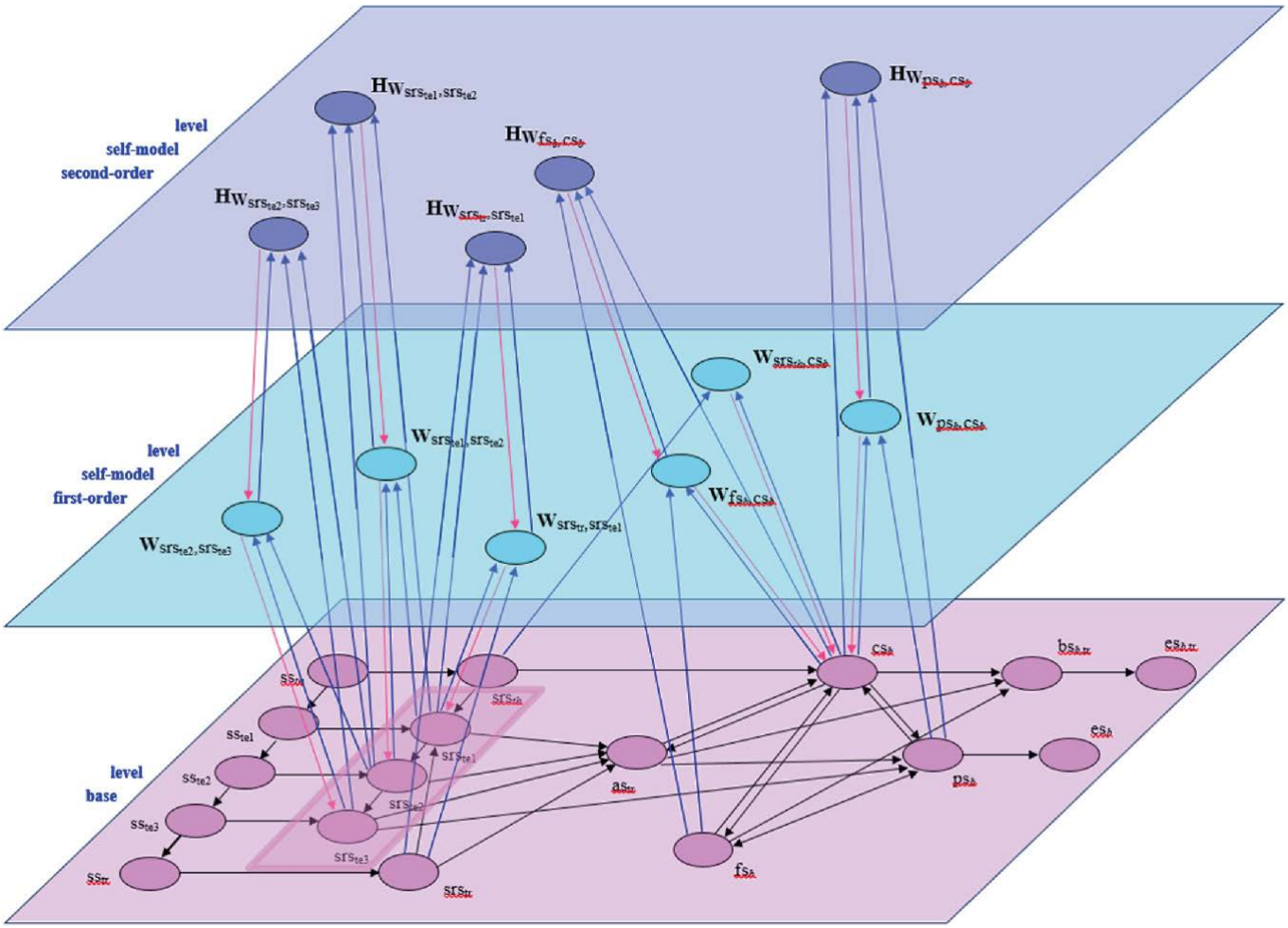


Fig. 5.1 Connectivity of the introduced second-order adaptive network model with the developed mental model for the traumatic course of events highlighted

1.
The Base Level

This level includes all *basic* (non-adaptive/non-learning) *processes* of the conceptual model.

2.
The First-Order Self-Model Level (or First Reification Level)

On this level, states are added that represent (adaptive) network characteristics of the base level. For example, a *self-model state* $\mathbf{W}_{X,Y}$ can be added to represent an adaptive connection weight $\boldsymbol{\omega}_{X,Y}$, or a *self-model state* $\mathbf{H}_Y$ can be added to represent a speed factor $\boldsymbol{\eta}_Y$. In the model in this way the learning of several connections in the base level takes place through Hebbian learning. These learning connections are represented by the dynamics of the **W**-states in the blue middle plane. This first-order self-model enables adaptation of the connections of the mental model in the base level.

3. The Second-Order Self-Model Level (or Second Reification Level)

Because the learning itself is adaptive as well, another level is added on top of the first-order self-model level: the second-order self-model level. This level allows to *control the learning speed* of the states $\mathbf{W}_{X,Y}$ for the learning connections by adding state $\mathbf{H}_{\mathbf{W}_{X,Y}}$ here representing the speed factor of $\mathbf{W}_{X,Y}$.

See for the connectivity of the network model Fig. 5.1; Table 5.1 shows the states and brief explanations of them. Within the network model, the first-order adaptation based on the Hebbian learning principle has been modeled by using a *connectivity self-model* (in the blue plane) based on self-model states $\mathbf{W}_{X,Y}$ representing connection weights $\boldsymbol{\omega}_{X,Y}$. These self-model states need incoming and outgoing connections to let them function within the network. To incorporate the 'firing together' part of (2) from Sect. 5.2, for the self-model's connectivity, incoming connections from X and Y to $\mathbf{W}_{X,Y}$ are used; see Fig. 5.1 (upward arrows in blue). These upward connections have weight 1. Also a connection from $\mathbf{W}_{X,Y}$ to itself with weight 1 is used to model persistence of the learnt effect; in pictures they are usually left out. In addition, an outgoing connection from $\mathbf{W}_{X,Y}$ to state Y is used to indicate where this self-model state $\mathbf{W}_{X,Y}$ has its effect; again see Fig. 5.1 (pink downward arrow). The downward connection indicates that the value of $\mathbf{W}_{X,Y}$ is actually used for the connection weight of the connection from X to Y. For the *aggregation characteristics* of the first-order self-model, the Hebbian learning rule is defined by the combination function $\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W)$ for self-model state $\mathbf{W}_{X,Y}$ from Table 5.4.

Table 5.1 The states in the network model and their explanation

state	explanation
X_1 ss_{te1}	Sensor state for traumatic event phase 1: observation te1
X_2 ss_{te2}	Sensor state for traumatic event phase 2: observation of action te2
X_3 ss_{te3}	Sensor state for traumatic event phase 3: observation of effect te3
X_4 ss_{tr}	Sensor state for trigger tr for the traumatic event sequence te
X_5 ss_{th}	Sensor state for input from therapy *th*
X_6 srs_{te1}	Sensory representation state for traumatic event phase 1: observation te1
X_7 srs_{te2}	Sensory representation state for traumatic event phase 2: action te2
X_8 srs_{te3}	Sensory representation state for traumatic event phase 3: effect te3
X_9 srs_{tr}	Sensory representation state for trigger tr for traumatic event sequence
X_{10} srs_{th}	Sensory representation state for therapy *th*
X_{11} as_{te}	Awareness state for the traumatic event te
X_{12} ps_b	Preparation state for emotional response *b*
X_{13} fs_b	Feeling state for emotional response *b*
X_{14} cs_b	Control state for emotional response *b*
X_{15} $bs_{b,te}$	Belief that emotional response *b* is from traumatic event te
X_{16} es_b	Bodily expressed emotional response *b*
X_{17} $es_{b,te}$	Expressing that emotional response *b* is from traumatic event te
X_{18} $\mathbf{W}_{srs_{te1},srs_{te2}}$	Representation state for weight of the connection from srs_{te1} to srs_{te2} for imprinting traumatic sequence
X_{19} $\mathbf{W}_{srs_{te2},srs_{te3}}$	Representation state for weight of the connection from srs_{te2} to srs_{te3} for imprinting traumatic sequence
X_{20} $\mathbf{W}_{srs_{tr},srs_{te1}}$	Representation state for weight of the connection from srs_{tr} to srs_{te1} for sensory preconditioning to link trigger tr to the traumatic sequence
X_{21} $\mathbf{W}_{ps_b,cs_b}$	Representation state for weight of the connection from ps_b to cs_b for learning of emotion regulation
X_{22} $\mathbf{W}_{fs_b,cs_b}$	Representation state for weight of the connection from fs_b to cs_b for learning of emotion regulation
X_{23} $\mathbf{W}_{th,cs_b}$	Representation state for weight of the connection from *th* to cs_b for learning of emotion regulation from therapy
X_{24} $\mathbf{Hw}_{srs_{te1},srs_{te2}}$	Control state for adaptation speed for weight of connection from srs_{te1} to srs_{te2}
X_{25} $\mathbf{Hw}_{srs_{te2},srs_{te3}}$	Control state for adaptation speed for weight of connection from srs_{te2} to srs_{te3}
X_{26} $\mathbf{Hw}_{srs_{tr},srs_{te1}}$	Control state for adaptation speed for weight of connection from srs_{tr} to srs_{te1}
X_{27} $\mathbf{Hw}_{ps_b,cs_b}$	Control state for adaptation speed for weight of connection from ps_b to cs_b
X_{28} $\mathbf{Hw}_{fs_b,cs_b}$	Control state for adaptation speed for weight of connection from fs_b to cs_b

The sensing of an example of a traumatic event te in the form of a sequence of steps is modeled by the sensor states ss_{te1}, ss_{te2}, ss_{te3}. For example, te1 (or traumatic event phase 1), is a potentially dangerous situation for a child you observe, 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 course of affairs, sensory representations srs_{te1}, srs_{te2}, srs_{te3} are activated for these phases te1, te2 and

te3, and by sensory preconditioning the connections between these sensory representations are learned. By this observational learning process, the mental model of the traumatic event 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. 5.1) with first-order self-model states $\mathbf{W}_{\text{srs}_{\text{te1}},\text{srs}_{\text{te2}}}$ and $\mathbf{W}_{\text{srs}_{\text{te1}},\text{srs}_{\text{te2}}}$. Similarly, the connection between the sensory representations of the trigger tr and the traumatic event sequence is learnt based on sensory preconditioning, represented by $\mathbf{W}_{\text{srs}_{\text{tr}},\text{srs}_{\text{te1}}}$. These newly formed connections activate the mental model as a form of internal mental simulation, every time the trigger is sensed. For the traumatized person this shows as an internal flashback movie of the traumatic sequence. In turn, this flashback movie activates the related negative emotions experienced at the original traumatic event.

In contrast to what was believed earlier, such learnt connections usually do not show any form of natural extinction; e.g., (Levin and Nielsen 2007), p. 507. Therefore, to make their effect more bearable, the only option is to suppress the emotional consequences related to the trauma by activating the emotion regulation control state cs_b. However, due to the high negative emotion levels the learning process for the activation of cs_b is impaired: learning speeds $\mathbf{H}_{\mathbf{W}_{\text{ps}_b,\text{cs}_b}}$ and $\mathbf{H}_{\mathbf{W}_{\text{fs}_b,\text{cs}_b}}$ are very low. Therefore, without any additional help the situation will stay as it is. But, following (Garcia 2002) the therapy *th* is able to temporarily reduce the level of negative emotions, so that $\mathbf{H}_{\mathbf{W}_{\text{ps}_b,\text{cs}_b}}$ and $\mathbf{H}_{\mathbf{W}_{\text{fs}_b,\text{cs}_b}}$ get higher values. Due to this, learning of the connections to the control state takes place: $\mathbf{W}_{\text{ps}_b,\text{cs}_b}$ and $\mathbf{W}_{\text{fs}_b,\text{cs}_b}$ get higher values.

5.3.3 Transcribing the Conceptual Model Into Role Matrices

To allow for easy formalization of the conceptual model into role matrices and an executable computational model, we use generic ways to describe the states, intra-level connections and interlevel connections. See an abstracted overview of all types of states and connections used in the model in Tables 5.2 and 5.3.

Table 5.2 Overview of types of states

State name	Representation
ss_y	Sensor state for state *y* in the world
srs_y	Sensory representation state for *y*
as_s	Awareness state for *s*
fs_b	Feeling state for feeling *b*
ps_b	Preparation state for feeling *b*

State name	Representation
cs_b	Control state for feeling b
bs_b	Belief state for feeling b
es_b	Execution state for feeling b
$W_{X,Y}$	Connection weight representation state for connection $X \rightarrow Y$
$H_{W_{X,Y}}$	Learning control state for the connection weight state for connection $X \rightarrow Y$

Table 5.3 Overview of types of connections

Connection	Representation	Connection Type
$X \rightarrow Y$	Connection between base states X and Y	Intra-level (horizontal) connection
$X \rightarrow \mathbf{W}_{X,Y}$ $Y \rightarrow \mathbf{W}_{X,Y}$	Connections from base level states X and Y to connection adaptation state $\mathbf{W}_{X,Y}$ to support the Hebbian learning formation	Interlevel connection, upward from the base level to the first-order self-model level
$\mathbf{W}_{X,Y} \rightarrow Y$	Connection from connection adaptation state $\mathbf{W}_{X,Y}$ to base state Y; these connections effectuate the learnt connection	Interlevel connection, downward from the first-order self-model level to the base level
$\mathbf{W}_{X,Y} \rightarrow H_{W_{X,Y}}$ $X \rightarrow H_{W_{X,Y}}$ $Y \rightarrow H_{W_{X,Y}}$	Connections from connection adaptation state $\mathbf{W}_{X,Y}$ and base level states X and Y to learning control state $H_{W_{X,Y}}$	Interlevel connections, upward from the base level to the second-order self-model level, and upward from the first-order self-model level to the second-order self-model level
$H_{W_{X,Y}} \rightarrow W_{X,Y}$	Connection from learning control state $H_{W_{X,Y}}$ to adaptive connection adaptation state $\mathbf{W}_{X,Y}$ to effectuate learning control	Interlevel connection, downward from the third level to the second level

The model with connectivity shown in Fig. 5.1 was then specified by tables in role matrix format: Connectivity characteristics (1), aggregation characteristics (2) and timing characteristics (3); see Sect. 5.6. Four different combination functions from the library are used that each serve a different purpose; see Table 5.4.

Table 5.4 The combination functions used from the library

Combination function	Notation	Formula	Parameters
Advanced logistic sum	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, \ldots, V_k)$	$\left[\frac{1}{1+e^{-\sigma(V_1+\cdots+V_k-\tau)}} - \frac{1}{1+e^{\sigma\tau}}\right](1+e^{-\sigma\tau})$	Steepness $\sigma > 0$ Excitability threshold V

Combination function	Notation	Formula	Parameters
Hebbian learning	**hebb**$_{\mu}$(V_1, V_2, W)	$V_1 V_2 (1 - W) + \mu W$	Persistence factor $\gamma \geq 0$
Steponce	**steponce**(V)	1 if $\alpha \leq t \leq \beta$, else 0	$\alpha \geq 0$ begin, $\beta \geq \alpha$ end time
Stepmod	**stepmod** ρ, δ $(V_1, \ldots, V_k)$	0 if $t \mod \rho < \delta$, else 1	Repetition $\rho \geq 0$ Duration $\rho \geq 0$

The advanced logistic sum combination function combines influences of multiple states by adding them but makes sure they stay between 0 and 1, with parameters steepness σ and threshold τ. The Hebbian learning combination function is used for learning of a connection weight. The stepmod function allows for an activation of states with a predefined length and frequency (here, that is used for the recurring trigger state). The steponce function allows for the activation of states with predefined length and start time (here, that is used for the therapy and trauma states).

5.4 Example Simulations

The role matrices listed in the Appendix Sect. 5.6 can easily be transferred to the dedicated software environment for simulations. Running the software loops over a chosen time period (in this case a time interval from 0 to 1400 with step size $\triangle t$ = 0.5) and provides as output a simulation graph for the model. In Fig. 5.2 the development of PTSD is shown based on traumatic event phases te1 to te3 in time period from 100 to 200 without applying therapy. The trigger also occurs from 100 to 200 and after that regularly recurs from 300 to 400, from 500 to 600, et cetera. In Fig. 5.3 the same is shown but this time therapy is taking place from time 400 to time 800 where the therapy leads to recovery. In both Figs. 5.2 and 5.3 in the time period from 100 to 200 the traumatic event sequence te1 to te3 in the world are sensed (via sensor states ss_{te1}, ss_{te2}, ss_{te3}) of which internal representations srs_{te1}, srs_{te2}, srs_{te3} are made. Due to sensory preconditioning (first-order adaptation based on Hebbian learning), the connections between them are developed (thus forming a mental model of the traumatic event sequence) and also a connection from the trigger representation srs_{tr} to srs_{te1}. Moreover, they trigger the negative emotional response preparation ps_b and feeling state fs_b, and these in turn reduce the adaptation speed (represented by the **H**-states) of the learning of the connections to the control state cs_b (second-order adaptation for metaplasticity). Therefore, no strengthening of the emotion regulation takes place, what would be needed to get rid of the negative feelings.

Every time period that the trigger recurs, due to the connection from srs_{tr} to srs_{te1} and the connections between srs_{te1}, srs_{te2}, srs_{te3}, the flasback movie is replayed (as a form of internal simulation of the mental model) and because of that the negative emotion and feeling are activated to high values again.

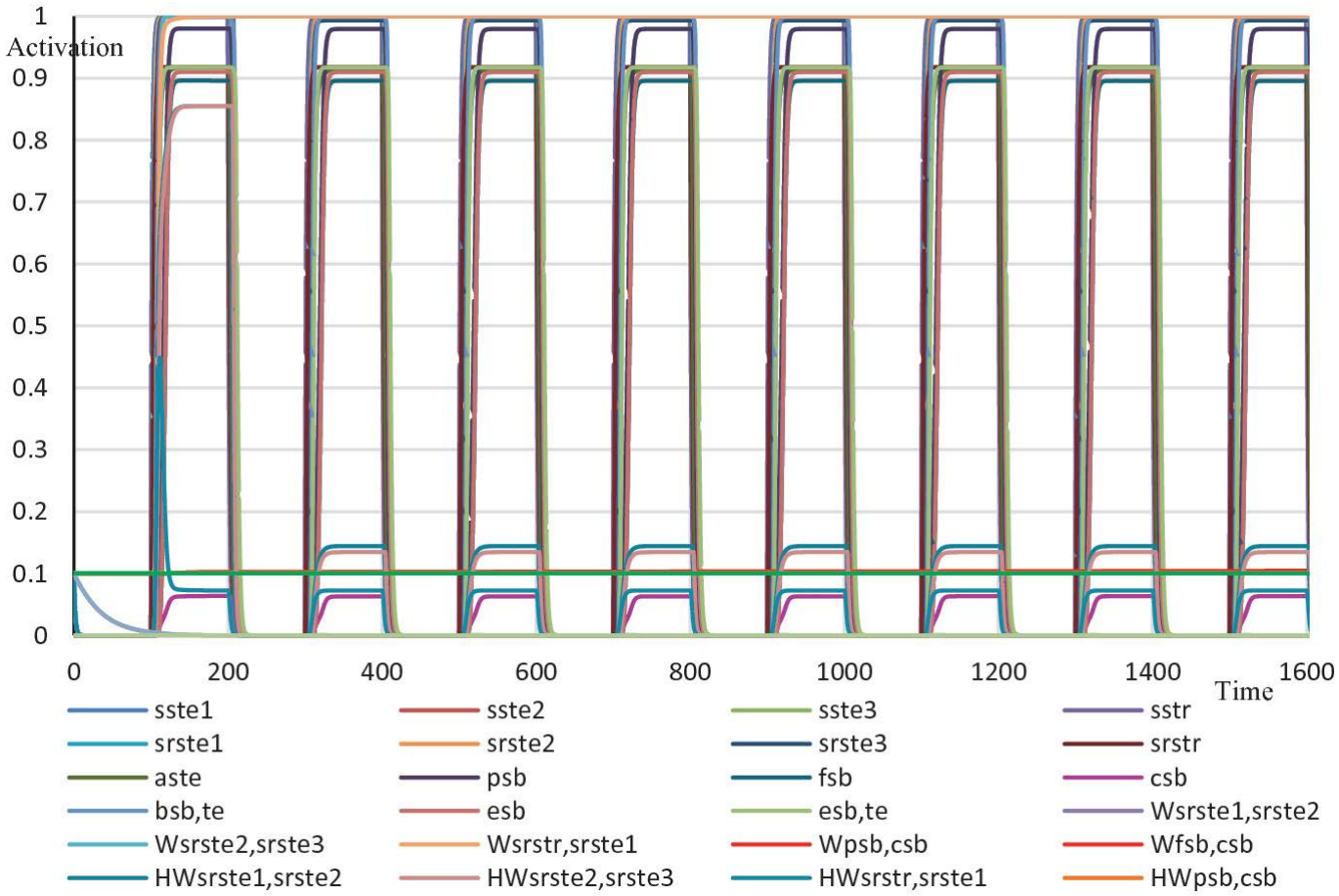


Fig. 5.2 Development of PTSD without using therapy. The trauma develops from time 100 to 200. The trigger also occurs from 100 to 200 and after that regularly recurs from 300 to 400, from 500 to 600, et cetera. No recovery from PTSD takes place

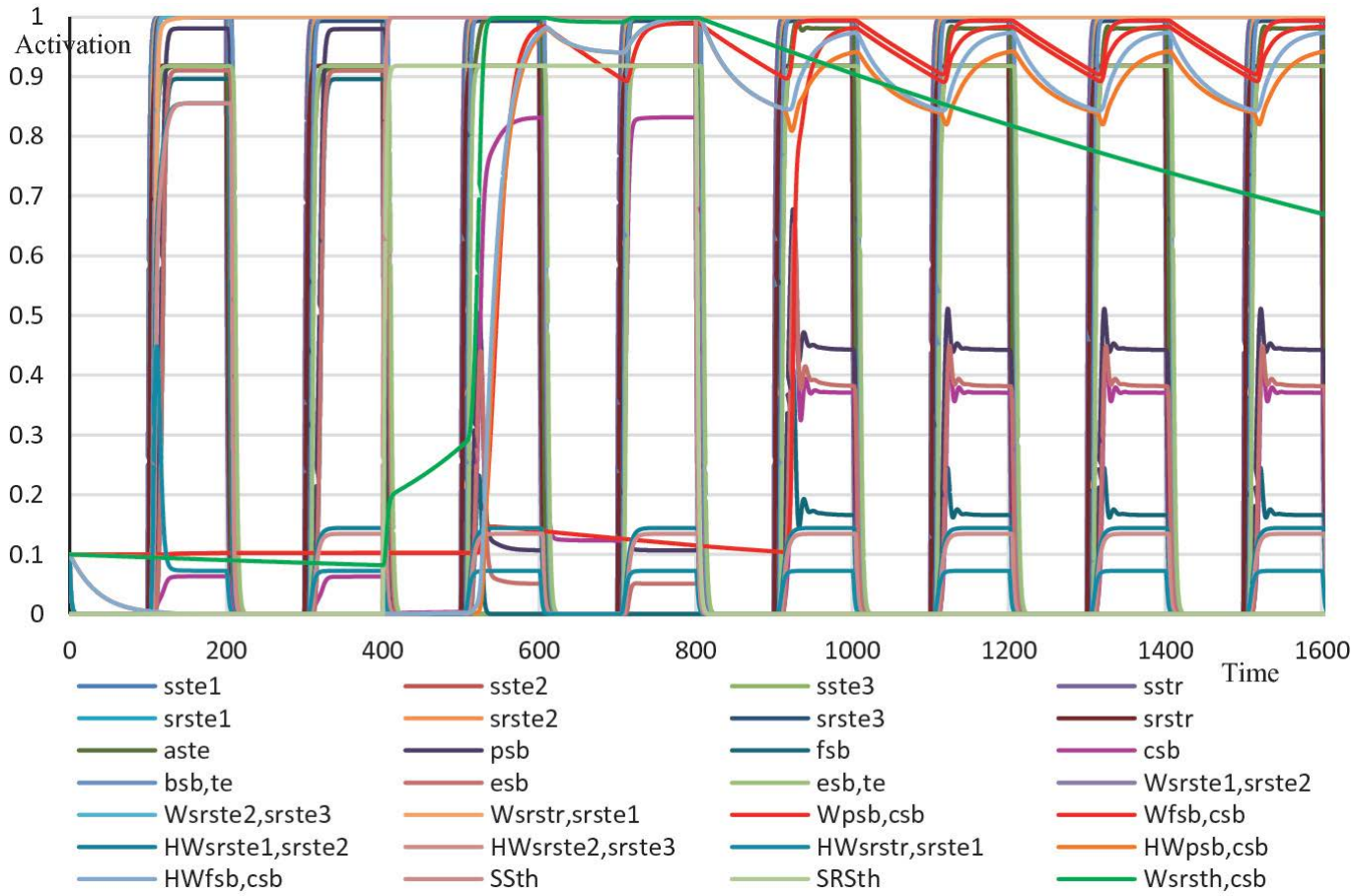


Fig. 5.3 Development of PTSD and recovery using therapy. Again, the trauma develops from time 100 to 200 and the trigger also occurs from 100 to 200 and after that regularly recurs from 300 to 400, from 500 to 600, et cetera. In this case therapy takes place from time 400 to 800 which leads to recovery

In Fig. 5.3 it is shown how the therapy temporarily (from time 400 to 800) reduces the negative emotion and feeling due to which the adaptation speeds (represented by the **H**-states) for the connections to control state cs_b increase again (second-order adaptation for metaplasticity) and therefore the learning of these connections to cs_b takes place. This finally results in much lower activations of the emotion and feeling states due to the strengthened emotion regulation.

5.5 Discussion

In this work, a second-order adaptive model was developed to allow for simulation of the formation of a mental model of a trauma that is built up over time and its emotional responses, and neurological processes of how a stimulus can become a trigger to activate this mental model. Furthermore, the influence of therapy on the ability of an individual to control the emotional response to the trauma mental model was explored. Most of the material was adopted from (Van

Ments and Treur 2021). The computational model was developed following the approach described in (Treur 2020b), using the following steps:

- A conceptual causal network model was designed based on literature on patients with PTSD and existing theories and models about PTSD and emotion regulation
- The conceptual causal network model was translated into role matrices format
- The role matrices were used in the dedicated software environment to obtain simulations; this software environment is available at https://www.researchgate.net/project/Network-Oriented-Modeling-Software.

Different simulation experiments were done, for individuals developing a trigger response, individuals not developing a trigger response, and individuals receiving therapy.

Other work addressing computational modelling for trauma development and recovery can be found in (Formolo et al. 2017; Naze and Treur 2011; Naze and Treur 2012). However, none of these previous works allowed for the adaptation of the learnt connections of the mental model and therapy. In addition, in (Naze and Treur 2011,2012) it is assumed that already built-in upward connections for the emotion regulation exist and are static, while in the model presented here an important part of the development of a trauma is the learning for the mental model of the traumatic course of affairs. In another comparison, (Formolo et al. 2017) addresses social support instead of the type of therapy suggested by Garcia (2002) and used in the current chapter. Moreover, the underlying second-order adaptation process as explained extensively by (Garcia 2002) is fully addressed here while it is ignored in (Formolo et al. 2017; Naze and Treur 2011; Naze and Treur 2012). Finally, in the current chapter the source of the trauma can be a process taking place over a longer time period with a successive course of events over time, and modeled in the form of an internal mental model that can be replayed as a flashback movie, while in (Formolo et al. 2017; Naze and Treur 2011; Naze and Treur 2012) only one traumatic state at one time point is assumed where a flashback is only one static image, which is not quite realistic.

The second-order adaptive model described in this chapter can be used as a basis for development of integrated computing applications to support PTSD therapy or to develop virtual characters illustrating the processes involved in patients with PTSD. In such contexts, also possibilities may be exploited for further validation of the model.

5.6 Appendix: Full Specification of the Adaptive Network Model

In this section, the full specification of the model is provided in terms of role matrices, which is the standard format used for design of a model and which can also be used as input for the software environment. Each role matrix has 3 color blocks that match the colors in the graphical representation used earlier: pink for the base level, blue for the first-order self-model states and purple for the second-order self-model states.

Connectivity characteristics role matrices.

The first role matrix **mb** (see Fig. 5.4, left hand side) represents the base connections between all the states, as presented in the graphical representation in Fig. 5.1. For example, row 6 with state srs_{te1} (also indicated by X_6) has two incoming connection: one from ss_{te1} and one from srs_{tr}. Role matrix **mcw** (see Fig. 5.4, right hand side) shows the weights **ω** of the connections presented in role matrix **mb**. There is a difference in nonadaptive (green) and adaptive connections (peach-red). For example, again row 6 with state srs_{te1} with two incoming connections: one non adaptive from ss_{te1} with connection weight 1 and one adaptive connection from srs_{tr}, presented by state X_{20}, i.e. $\mathbf{W}_{srs_{tr},srs_{te1}}$ (a first-order self-model state).

mb base connectivity		1	2	3	4	5	mcw connection weights		1	2	3	4	5
X_1	ss_{te1}	X_1					X_1	ss_{te1}	1				
X_2	ss_{te2}	X_1					X_2	ss_{te2}	1				
X_3	ss_{te3}	X_2					X_3	ss_{te3}	1				
X_4	ss_{tr}	X_4					X_4	ss_{tr}	1				
X_5	ss_{th}	X_5					X_5	ss_{th}	1				
X_6	srs_{te1}	X_1	X_9				X_6	srs_{te1}	1	X_{20}			
X_7	srs_{te2}	X_2	X_6				X_7	srs_{te2}	1	X_{18}			
X_8	srs_{te3}	X_3	X_7				X_8	srs_{te3}	1	X_{19}			
X_9	srs_{tr}	X_4					X_9	srs_{tr}	1				
X_{10}	srs_{th}	X_5					X_{10}	srs_{th}	1				
X_{11}	as_{te}	X_6	X_7	X_8	X_9	X_{14}	X_{11}	as_{te}	1	1	1	1	1
X_{12}	ps_b	X_8	X_{11}	X_{13}	X_{14}		X_{12}	ps_b	1	1	-0.5	1	
X_{13}	fs_b	X_{12}	X_{14}				X_{13}	fs_b	1	-0.5			
X_{14}	cs_b	X_8	X_{11}	$\mathbf{X_{10}}$	$\mathbf{X_{12}}$	$\mathbf{X_{13}}$	X_{14}	cs_b	0.3	0.3	X_{23}	X_{21}	X_{22}
X_{15}	$bs_{b,te}$	X_8	X_{11}	X_{13}			X_{15}	$bs_{b,te}$	1	1			
X_{16}	es_b	X_{12}					X_{16}	es_b	1				
X_{17}	$es_{b,te}$	X_{15}					X_{17}	$es_{b,te}$	1				
X_{18}	$\mathbf{W}_{srs_{te1},srs_{te2}}$	X_6	X_7	X_{18}			X_{18}	$\mathbf{W}_{srs_{te1},srs_{te2}}$	1	1	1		
X_{19}	$\mathbf{W}_{srs_{te2},srs_{te3}}$	X_7	X_8	X_{19}			X_{19}	$\mathbf{W}_{srs_{te2},srs_{te3}}$	1	1	1		
X_{20}	$\mathbf{W}_{srs_{tr},srs_{te1}}$	X_9	X_6	X_{20}			X_{20}	$\mathbf{W}_{srs_{tr},srs_{te1}}$	1	1	1		
X_{21}	$\mathbf{W}_{ps_b,cs_b}$	X_{12}	X_{14}	X_{21}			X_{21}	$\mathbf{W}_{ps_b,cs_b}$	1	1	1		
X_{22}	$\mathbf{W}_{fs_b,cs_b}$	X_{13}	X_{14}	X_{22}			X_{22}	$\mathbf{W}_{fs_b,cs_b}$	1	1	1		
X_{23}	$\mathbf{W}_{th,cs_b}$	X_{15}	X_{14}	X_{23}			X_{23}	$\mathbf{W}_{th,cs_b}$	1	1	1		
X_{24}	$\mathbf{Hw}_{srste1,srste2}$	X_7	X_8	X_{18}	X_{24}		X_{24}	$\mathbf{Hw}_{srste1,srste2}$	1	1	-0.5	1	
X_{25}	$\mathbf{Hw}_{srste2,srste3}$	X_6	X_7	X_{19}	X_{25}		X_{25}	$\mathbf{Hw}_{srste2,srste3}$	1	1	-0.5	1	
X_{26}	$\mathbf{Hw}_{srstr,srste1}$	X_9	X_6	X_{20}	X_{26}		X_{26}	$\mathbf{Hw}_{srstr,srste1}$	1	1	-0.5	1	
X_{27}	$\mathbf{Hw}_{psb,csb}$	X_{12}	X_{14}	X_{21}	X_{27}		X_{27}	$\mathbf{Hw}_{psb,csb}$	-0.5	1	0	1	
X_{28}	$\mathbf{Hw}_{fsb,csb}$	X_{13}	X_{14}	X_{22}	X_{28}		X_{28}	$\mathbf{Hw}_{fsb,csb}$	-0.5	1	0	1	

Fig. 5.4 Connectivity characteristics: role matrices **mb** and **mcw**

Aggregation characteristics role matrices.

In order to transform the presented graphical model with states and connections into a numerical model each state needs a combination function $\mathbf{c}_Y$**(..)** to aggregate the impacts of other states on state *Y*. Role matrix **mcfw** (see Fig. 5.5) specifies which combination function is used for each state. For example, state ss_{te1} uses the steponce function, so this is indicated with a combination function weight **γ** of 1 in the fourth column.

mcfw combination function weights		1 alogistic	2 hebb	3 stepmod	4 steponce
X_1	ss_{te1}				1
X_2	ss_{te2}	1			
X_3	ss_{te3}	1			
X_4	ss_{tr}			1	
X_5	ss_{th}				1
X_6	srs_{te1}	1			
X_7	srs_{te2}	1			
X_8	srs_{te3}	1			
X_9	srs_{th}	1			
X_{10}	srs_{tr}	1			
X_{11}	as_{te}	1			
X_{12}	ps_b	1			
X_{13}	fs_b	1			
X_{14}	cs_b	1			
X_{15}	$bs_{b,te}$	1			
X_{16}	es_b	1			
X_{17}	$es_{b,te}$	1			
X_{18}	$\mathbf{W}_{srs_{te1},srs_{te2}}$		1		
X_{19}	$\mathbf{W}_{srs_{te2},srs_{te3}}$		1		
X_{20}	$\mathbf{W}_{srs_{tr},srs_{te1}}$		1		
X_{21}	$\mathbf{W}_{ps_b,cs_b}$		1		
X_{22}	$\mathbf{W}_{fs_b,cs_b}$		1		
X_{23}	$\mathbf{W}_{th,cs_b}$		1		
X_{24}	$\mathbf{Hw}_{srste1,srste2}$	1			
X_{25}	$\mathbf{Hw}_{srste2,srste3}$	1			
X_{26}	$\mathbf{Hw}_{srstr,srste1}$	1			
X_{27}	$\mathbf{Hw}_{psb,csb}$	1			
X_{28}	$\mathbf{Hw}_{fsb,csb}$	1			

Fig. 5.5 Aggregation characteristics: role matrix **mcfw**

Role matrix **mcfp** (see Fig. 5.6) defines the exact parameters used for each state and function. The parameters used depend on the combination function.

mcfp	combination function parameters	1 alogistic σ	τ	2 hebb μ	3 stepmod ρ	δ	4 steponce α	β
X_1	ss_{te1}						100	200
X_2	ss_{te2}	20	0.5					
X_3	ss_{te3}	20	0.5					
X_4	ss_{tr}				200	100		
X_5	ss_{th}	5	0.5				400	1600
X_6	srs_{te1}	20	0.5					
X_7	srs_{te2}	20	0.5					
X_8	srs_{te3}	10	0.5					
X_9	srs_{tr}	5	0.5					
X_{10}	srs_{th}	5	0.5					
X_{11}	as_{te}	5	0.5					
X_{12}	ps_b	5	0.5					
X_{13}	fs_b	5	0.5					
X_{14}	cs_b	5	0.5					
X_{15}	$bs_{b,te}$	5	0.5					
X_{16}	es_b	5	0.5					
X_{17}	$es_{b,te}$	5	0.5					
X_{18}	$W_{srs_{te1},srs_{te2}}$			1				
X_{19}	$W_{srs_{te2},srs_{te3}}$			1				
X_{20}	$W_{srs_{tr},srs_{te1}}$			1				
X_{21}	W_{ps_b,cs_b}			0.999				
X_{22}	W_{fs_b,cs_b}			0.999				
X_{23}	W_{th,cs_b}			0.909				
X_{24}	$H_{W_{srste1},srste2}$	5	2					
X_{25}	$H_{W_{srste2},srste3}$	5	2					
X_{26}	$H_{W_{srstr},srste1}$	5	2					
X_{27}	$H_{W_{psb},csb}$	5	2					
X_{28}	$H_{W_{fsb},csb}$	5	2					

Fig. 5.6 Aggregation characteristics: role matrix **mcfp**

Timing characteristics role matrix and initial values.

The speed factors $\boldsymbol{\eta}_Y$ are specified in role matrix **ms;** see Fig. 5.7. Concerning the initial values, the simulated scenario starts with most state values 0. The learning states have initial value >0, chosen at 0.1.

ms speed factors		1	initial values		
X_1	ss_{te1}	0.5	X_1	ss_{te1}	0
X_2	ss_{te2}	0.5	X_2	ss_{te2}	0
X_3	ss_{te3}	0.5	X_3	ss_{te3}	0
X_4	ss_{tr}	0.5	X_4	ss_{tr}	0
X_5	ss_{th}	0.5	X_5	ss_{th}	0
X_6	srs_{te1}	0.5	X_6	srs_{te1}	0
X_7	srs_{te2}	0.5	X_7	srs_{te2}	0
X_8	srs_{te3}	0.5	X_8	srs_{te3}	0
X_9	srs_{tr}	0.5	X_9	srs_{tr}	0
X_{10}	srs_{th}	0.5	X_{10}	srs_{th}	0
X_{11}	as_{te}	0.5	X_{11}	as_{te}	0
X_{12}	ps_b	0.5	X_{12}	ps_b	0
X_{13}	fs_b	0.5	X_{13}	fs_b	0
X_{14}	cs_b	0.5	X_{14}	cs_b	0
X_{15}	$bs_{b,te}$	0.5	X_{15}	$bs_{b,te}$	0
X_{16}	es_b	0.5	X_{16}	es_b	0
X_{17}	$es_{b,te}$	0.5	X_{17}	$es_{b,te}$	0
X_{18}	$W_{srs_{te1},srs_{te2}}$	X_{24}	X_{18}	$W_{srs_{te1},srs_{te2}}$	0.1
X_{19}	$W_{srs_{te2},srs_{te3}}$	X_{25}	X_{19}	$W_{srs_{te2},srs_{te3}}$	0.1
X_{20}	$W_{srs_{tr},srs_{te1}}$	X_{26}	X_{20}	$W_{srs_{tr},srs_{te1}}$	0.1
X_{21}	W_{ps_b,cs_b}	X_{27}	X_{21}	W_{ps_b,cs_b}	0.1
X_{22}	W_{fs_b,cs_b}	X_{28}	X_{22}	W_{fs_b,cs_b}	0.1
X_{23}	W_{th_b,cs_b}	0.5	X_{23}	W_{th_b,cs_b}	0.1
X_{24}	$Hw_{srste1,srste2}$	0.5	X_{24}	$Hw_{srste1,srste2}$	0.1
X_{25}	$Hw_{srste2,srste3}$	0.5	X_{25}	$Hw_{srste2,srste3}$	0.1
X_{26}	$Hw_{srstr,srste1}$	0.5	X_{26}	$Hw_{srstr,srste1}$	0.1
X_{27}	$Hw_{psb,csb}$	0.05	X_{27}	$Hw_{psb,csb}$	0.1
X_{28}	$Hw_{fsb,csb}$	0.05	X_{28}	$Hw_{fsb,csb}$	0.1

Fig. 5.7 Timing characteristics: role matrix **ms** and initial values

References

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

(1996)
[Crossref]

Admon, R., Milad, M.R., Hendler, T.: A causal model of post-traumatic stress disorder: disentangling predisposed from acquired neural abnormalities. Trends Cogn. Sci. **17**(7), 337–347 (2013)
[Crossref]

Akiki, T.J., Averill, C.L., Abdallah, C.G.: A Network-based neurobiological model of PTSD: evidence from structural and functional neuroimaging studies. Curr. Psychiatry. Rep. **19**, 81 (2017). https://doi.org/10.1007/s11920-017-0840-4
[Crossref]

Benbassat, J.: Role modeling in medical education: the importance of a reflective imitation. Acad. Med. **89**(4), 550–554 (2014)
[Crossref]

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

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

Duvarci, S., Pare, D.: Amygdala microcircuits controlling learned fear. Neuron **82**, 966–980 (2014)
[Crossref]

Formolo, D., Van Ments, L., Treur, J.: A computational model to simulate development and recovery of traumatised patients. Biol. Inspired Cognitive Architect. **21**, 26–36 (2017)
[Crossref]

Fitzgerald, J.M., DiGangi, J.A., Phan, K.L.: Functional neuroanatomy of emotion and its regulation in PTSD. Harv Rev Psychiatry **26**(3), 116–128 (2018)
[Crossref]

Garcia, R.: Stress, metaplasticity, and antidepressants. Curr. Mol. Med. **2**, 629–638 (2002)
[Crossref]

Hall, G.: Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. Anim. Learn. Behav. **24**, 233–255 (1996)
[Crossref]

Hebb, D.O.: The organization of behavior: A neuropsychological theory. Wiley (1949)

Holmes, S.E., Scheinost, D., DellaGioia, N., Davis, M.T., Matuskey, D., Pietrzak, R.H., Hampson, M., Krystal, J.H., Esterlis, I.: Cerebellar and prefrontal cortical alterations in PTSD: structural and functional evidence. Chronic Stress **2**, 1–11 (2018). https://doi.org/10.1177/2470547018786390
[Crossref]

Keysers, C., Gazzola, V.: Hebbian learning and predictive mirror neurons for actions, sensations and emotions. Philos Trans. R Soc. Lond B Biol. Sci. **369**, 20130175 (2014)
[Crossref]

Levin, R., Nielsen, T.A.: Disturbed dreaming, posttraumatic stress disorder, and affect distress: a review and neurocognitive model. Psychol. Bull. **133**, 482–528 (2007)
[Crossref]

Naze, S., Treur, J.: A computational agent model for post-traumatic stress disorders. In: Samsonovich, A.V.,

Johannsdottir, K.R. (eds.), Proceedings of the Second International Conference on Biologically Inspired Cognitive Architectures, BICA'11, pp. 249–261. IOS Press (2011)

Naze, S., Treur, J.: A computational agent model for development of posttraumatic stress disorders by Hebbian learning. In: T. Huang et al. (eds.), Proceedings of the 19th International Conference on Neural Information Processing, ICONIP'12, Part II. Lecture Notes in Computer Science, vol. 7664, pp. 141–151. Berlin Heidelberg, Springer (2012)

Ochsner, K.N., Gross, J.J.: The neural bases of emotion and emotion regulation: A valuation perspective. Handbook of emotional regulation (2nd ed.), pp. 23–41. Guilford, New York (2014)

Panksepp, J., Biven, L.: The archaeology of mind: Neuroevolutionary origins of human emotions. New York, W.W. Norton (Ch 1). (2012)

Parsons, R.G., Ressler, K.J.: Implications of memory modulation for post-traumatic stress and fear disorders. Nat. Neurosci. **16**(2), 146–153 (2013)
[Crossref]

Shatz, C.J.: The developing brain. Sci. Am. **267**, 60–67 (1992). https://doi.org/10.1038/scientificamerican0992-60

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions. Springer Publishers, Cham Switzerland (2016).

Treur, J.: Modeling higher-order adaptivity of a network by multilevel network reification. Netw. Sci. **8**, S110–S144 (2020)
[Crossref]

Treur, J.: Network-oriented modeling for adaptive networks: Designing Higher-order Adaptive Biological, Mental and Social Network Models. Springer Nature, Cham Switzerland (2020b)

Van Gog, T., Paas, F., Marcus, N., Ayres, P., Sweller, J.: The mirror neuron system and observational learning: implications for the effectiveness of dynamic visualizations. Educ. Psychol. Rev. **21**(1), 21–30 (2009)
[Crossref]

Van Ments, L., Treur, J.: A Higher-order adaptive network model to simulate development of and recovery from PTSD. In: Proc. of the 21th International Conference on Computational Science, ICCS'21, pp. 154–166. Lecture Notes in Computer Science, vol. 12743. Springer Nature Switzerland (2021)

Webb, T.L., Miles, E., Sheeran, P.: Dealing with feeling: a meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. Psychol. Bull. **138**(4), 775 (2012)
[Crossref]

Zandvakili, A., Barredo, J., Swearingen, H.R., Aiken, E.M., Berlow, Y.A., Greenberg, B.D., Carpenter, L.L., Philip, N.S.: Mapping PTSD symptoms to brain networks: a machine learning study. Transl Psychiatry **10**, e195 (2020)

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_6

6. 'What if I Would Have Done Otherwise…': A Controlled Adaptive Network Model for Mental Models in Counterfactual Thinking

Raj Bhalwankar[1] ✉ and Jan Treur[1] ✉
(1) Social AI Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
In this chapter counterfactual thinking is addressed based on literature mainly from Neuroscience and Psychology. A detailed literature review was conducted in identifying processes, neural correlates and theories related to counterfactual thinking from different disciplines. A familiar scenario with respect to counterfactual thinking was identified. Based on the literature, an adaptive self-modeling network model was designed. This model captures the complex process of counterfactual thinking, the mental models that are involved, and the learning and control.

Keywords Counterfactual thinking – Adaptive network model – Mental model – Learning – Control

6.1 Introduction

Human beings have a great ability to think and infer how a current situation (especially goal failure) could have turned out differently given a set of alternative actions or decisions they could have chosen from (Byrne 2002). This process of deconstructing the current reality to imagine (a) new one(s) is called counterfactual thinking (Timberlake 2019). Such a type of thinking is important as in the first place it helps in making sense of the past, in planning actions and in

making emotional and social judgements. Not less important, in the second place it plays an important functional role to guide adaptive behavior and learning (Van Hoeck, Watson, Barbey 2015) for the own benefit for the future. Such learning is a form of learning from mistakes, which involves the notion of regret which arises from comparing the alternative realities. This type of learning helps to generate new courses of actions which, after the failure experience, are believed to be more successful when similar situations occur in future. Various parts of the brain have been implicated to play a role in counterfactual thinking.

Yet also many questions about counterfactual thinking still have no full answers. How does the process of counterfactual thinking actually work in day-to-day life? What is its role in mental health, learning and decision-making? How does it affect our emotional health and how does it update our beliefs or perceptions? How can counterfactual thinking prove useful in AI applications like reinforcement learning?

Thus, the present study is meant to contribute some answers to these questions by providing a Neuroscience-inspired controlled adaptive network model that is able to simulate processes of counterfactual thinking, including the involved mental models, the learning effects of it, and control over it. Computational modeling plays an important role here in making sense of behavioral and neurological data. Computational models represent different 'algorithmic hypotheses' about how behavior is generated (Wilson and Collins 2019). Such simulations involve running the model with specific parameter settings to generate 'fake' behavioral data. These simulated data can then be analyzed in much the same way as one would analyze real data, to make precise, falsifiable predictions about qualitative and quantitative patterns in the data. These simulations contribute to theory building by making theoretical predictions more precise and testable (Wilson and Collins 2019).

Mental models are essential for construction of knowledge and play a crucial role in learning, retrieving and problem solving (Kim 2004; Rouse and Morris 1986; Olson 1992). Van Hoeck et al. (2015) proposed that counterfactuals depend upon the mental models of alternative possibilities in form of mental simulations, which suggests that modeling approaches to mental model development can be used to study counterfactual thinking. In the present study, a network-oriented modeling approach was utilized to study the process of counterfactual thinking based on literature and neuro-scientific evidence. Network-oriented modeling is a useful method to represent the complex real-world processes concerning human beings and has proven to be able to address adaptivity and control that play an important role in counterfactual thinking and the mental models involved.

The chapter begins in Sect. 6.2 with a brief literature overview of the existing state of research related to counterfactual thinking. Then, after a brief

introduction of the modeling approach used in Sect. 6.3, in Sect. 6.4 the design of the developed controlled adaptive network model with its various parts is discussed. Simulations of some example scenarios are discussed in Sect. 6.5; here it is shown that the model generates patterns as expected from the empirical literature. In Sect. 6.6, correctness of the implemented model against its conceptual design specifications is verified by analysis of stationary points. Section 6.7 addresses discussion and conclusions.

6.2 Literature Review

As stated earlier, counterfactual thinking can be helpful in learning from past mistakes and in developing more promising intentions for the future (Markman et al. 1993; Roese 1994; Sanna et al. 2002). Mental models of imagined past events or future outcomes that have not yet occurred support *counterfactual thinking* (Byrne 2002; Kahneman and Miller 1986). Norm Theory proposed by Kahneman and Miller (1986) provides a theoretical basis to describe the rationale for counterfactual thoughts. According to them, counterfactual thinking is driven by simulations of previously encoded exemplars and they emphasize the role of counterfactual thinking in reframing such scenarios—generating alternative possibilities that change the norms (and expectations) used to interpret a situation (Van Hoeck et al. 2015). The theory suggests that the counterfactual alternatives created depend on the ease of imagining different outcomes. The norms involve a pairwise comparison between a cognitive standard and an experiential outcome. A discrepancy that is created by such a comparison elicits an affective response which is influenced by the magnitude and direction of the difference.

Rational Imagination Theory proposed by Byrne (2005) says that the counterfactual imagination is rational and it depends on three assumptions: (1) humans are capable of rational thought; (2) they make inferences by thinking about possibilities; and (3) their counterfactual thoughts rely on thinking about possibilities, just as rational thoughts do. Byrne (2007) proposed a set of cognitive directives that guide these possibilities when people imagine alternatives. The theory states that individuals' ability to entertain multiple parallel models corresponding to alternative possibilities suggests that counterfactual thought is engaged to search the space of possible alternatives.

According to Byrne (2016), an algorithm to specify the mental representations and cognitive processes that create counterfactuals takes as input the relevant facts of the actual event and produces as output a counterfactual alternative. The intervening processes change aspects of the mental representation of the facts to create a second mental representation, the counterfactual alternative. According to (De Brigard et al. 2019), the dynamic

nature of memory reconstruction allows to mentally modify aspects of autobiographical memory when simulating on retrieval, leading to counterfactual thinking. Computational mechanisms underlying counterfactual thinking maintain and update two representations, the imagined alternative and the known or presupposed reality.

The neural representations of counterfactual inference that are implicated in the neural systems for constructing mental models of the past and future, involve prefrontal and medial temporal lobe structures of the brain (Fortin et al. 2002; Tulving and Markowitsch 1998). A functional perspective on counterfactual thinking views it as a useful, and essential component of behavior regulation. It considers counterfactual thoughts closely connected to goal cognitions where counterfactual thinking is activated usually by goal-failure (Epstude and Roese 2008). It suggests that at its root counterfactual thinking is a regulatory loop–governing behavior which operates through a negative feedback model. This model operates by preserving homeostasis by correcting behavior when a discrepancy is detected between the current state and an ideal reference state for example goal-progress. In the theory of core affect (Russell 2003), affective experiences function as indicators of a discrepancy between current and an ideal state, thus affect often mediates behavior change. Also, once the discrepancy reduces between the current state and the reference state, corrective activity is terminated. In a review on counterfactual thinking (Epstude and Roese 2008), it was mentioned that cognitive experiments indicated that counterfactual thinking influences behavior by either of two routes: (1) a content-specific pathway where specific informational events affect behavioral intentions which then influence the behavior itself and (2) a content neutral pathway which has indirect effects by affect, mindsets, and motivations.

The structured event complex theory proposed by Barbey et al. (2009), state that counterfactual thinking engages a network of regions within prefrontal cortex (PFC) that represent alternative goals, behavioral intentions, mindsets, motivations, and self-inferences that enable behavioral change and adaptation (Van Hoeck et al. 2015). They also stated that counterfactual thought depends on mental models of alternative possibilities that are represented in the form of structured event complexes (SEC). SEC is a goal-oriented set of events that is structured in sequence and represents event features (like agents, objects, actions, mental states, and background settings), social norms of behavior, ethical and moral rules and temporal event boundaries.

Building upon these frameworks, Van Hoeck et al. (2015) proposed that counterfactual thinking depends upon the co-ordination of multiple information processing systems that involve three neural networks: (1) the mental simulation network, (2) the cognitive control network, (3) the reward network.

Thus, they proposed three stages of processing in counterfactual thinking: Activation, Inference and Learning and Adaptation; see Fig. 6.1.

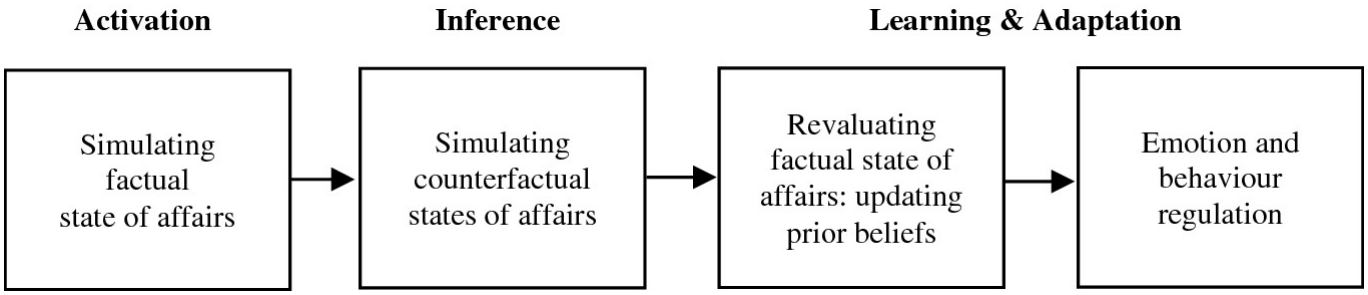


Fig. 6.1 Schematic overview of the stages in counterfactual thinking (Van Hoeck et al. 2015)

Activation. Counterfactual thoughts are triggered automatically in response to real world experiences, especially negative emotions triggered by violations of expectations and motivations (e.g., an unsuccessful application for a university), implicit/explicit goal failure (e.g., failing an exam), or close calls (e.g., missing the train by few seconds). As mentioned earlier, they proposed that counterfactuals depend upon the mental models of alternative possibilities in the form of mental simulations. Such simulations provide the foundation for constructing and evaluating mental models of reality and of imagined alternative possibilities. Van Hoeck et al. (2015) state that counterfactuals activate areas of the medial prefrontal cortex which are related to conflict detection.

Counterfactual Inference. Counterfactual inference concerns the internal simulation of mental models for alternative possibilities. It adheres to the 'nearest possible world' constraint. Thus, a counterfactual must closely model one's own experience of the real state of the world which helps in setting the constraint to specific situational features and prior knowledge of the situation. The alternatives suggested from such a constraint deviates only marginally from the reality, thus is probable. This also separates counterfactual thinking from less constrained forms of imagination or fantasy. For example, the chosen counterfactuals often are around factors like: (1) the factor that played the strongest role (2) the most deviation from expectation, or (3) is mostly under participant's own control. Apart from being influenced by meaning and relevance of specific events, counterfactuals are influenced by a specific individual perspective where implicit belief of attainability or self-efficacy plays an important role. The counterfactual outcome value plays an important role and impacts how an individual perceives the factual, as well as the experienced outcome and its relative value. Such evaluations of how a counterfactual could have been better lead to emotional and social reactions.

Learning and Adaptation. The information inferred by internal simulation based on a mental model for a counterfactual state of affairs will then get

incorporated in the representation of the current state of affairs leading to re-evaluation and updating of prior beliefs and action-values. This leads to lasting behavioral and affective modifications. Counterfactual thinking provides an opportunity to improve performance for the future and elicits behavioral motivations to pursue the counterfactual outcome. This regulates one's perception of control and preparedness, boosting persistence and performance (Van Hoeck et al. 2015).

6.3 The Modeling Approach for Controlled Adaptive Networks

Temporal-causal network models as addressed in (Treur 2020) can be represented by a conceptual representation and by a numerical representation. A conceptual representation involves representing in a declarative manner states and connections between them that represent the causal impacts of states on each other. The states are assumed to have activation levels that vary over time. Causal relations have weights. Furthermore, when more than one causal relation affects a state, some way to aggregate multiple causal impacts on a state is used. This aggregation is indicated by some combination function from a library. For the timing of the state dynamics, a speed factor is used, so that no synchronous processing is required. The notions for

- **connectivity** (*connection weights* $\boldsymbol{\omega}_{X,Y}$)
- **aggregation** (*combination functions* $\mathbf{c}_Y(..)$)
- **timing** (*speed factors* $\boldsymbol{\eta}_Y$)

are the network characteristics that define a conceptual representation of a temporal-causal network model; they are summarized in Table 6.1, in the first four rows.

Table 6.1 Conceptual and numerical representation of a temporal-causal network (Treur 2016)

Concepts	*Notation*	*Explanation*
States and connections	$X, Y, X \rightarrow Y$	Describes the nodes and links of a network structure
Connection weight	$\boldsymbol{\omega}_{X,Y}$	The connection weight $\boldsymbol{\omega}_{X,Y}$ represents the strength of the causal impact of state X on state Y with $X \rightarrow Y$
Aggregating multiple causal impacts	$\mathbf{c}_Y(..)$	For each state Y a combination function $\mathbf{c}_Y(..)$ is chosen to combine the causal impacts of other states on state Y

Concepts	*Notation*	*Explanation*
Timing of the causal effect	$\boldsymbol{\eta}_Y$	For each state Y a speed factor $\boldsymbol{\eta}_Y \geq 0$ is used to represent how fast a state is changing upon causal impact
Concepts	*Numerical representation*	*Explanation*
State values over time t	$Y(t)$	At each time point t each state Y in the model has a real number value
Single causal impact	$\mathbf{impact}_{X,Y}(t) = \boldsymbol{\omega}_{X,Y} X(t)$	At t state X with a connection to state Y has an impact $\boldsymbol{\omega}_{X,Y} X(t)$ on Y, using weight $\boldsymbol{\omega}_{X,Y}$
Aggregating multiple causal impacts	$\mathbf{aggimpact}_Y(t) = \mathbf{c}_Y(\mathbf{impact}_{X_1,Y}(t),\ldots,\mathbf{impact}_{X_k,Y}(t)) = \mathbf{c}_Y(\boldsymbol{\omega}_{X_1,Y} X_1(t), \ldots, \boldsymbol{\omega}_{X_1,Y} X_k(t))$	The aggregated impact of X_i on Y at t, is determined by applying combination function $\mathbf{c}_Y(..)$ on $\boldsymbol{\omega}_{X_i,Y} X_i(t)(t)$
Timing of the causal effect per state	$Y(t+\Delta t) = Y(t) + \boldsymbol{\eta}_Y[\mathbf{aggimpact}_Y(t) - Y(t)]\Delta t = Y(t) + \boldsymbol{\eta}_Y[\mathbf{c}_Y(\boldsymbol{\omega}_{X_1,Y} X_1(t), \ldots, \boldsymbol{\omega}_{X_1,Y} X_k(t)) - Y(t)]\Delta t$	The causal impact on Y is exerted over time gradually, using speed factor $\boldsymbol{\eta}_Y$

Combination functions can be selected from an available combination function library provided by the dedicated software environment that has been developed. For each state Y one or more basic combination functions $\mathbf{c}_j(..)$, $j = 1,..,m$ can be selected by indicating *combination function weights* $\boldsymbol{\gamma}_{j,Y}$ (real numbers) which makes that within the software environment a weighted average of these functions $\mathbf{c}_j(..)$, $j = 1,..,m$, from the library is used as combination function $\mathbf{c}_Y(..)$ for state Y; these basic combination functions $\mathbf{c}_j(..)$ have *combination function parameters* $\boldsymbol{\pi}_{i,j,Y}$. Currently there are more than 50 combination functions in the library. New combination functions can be added easily. The library has also facilities to apply function composition to define new functions by composing any number of functions from the library. In the model presented here, for the states, the two basic combination functions shown in Table 6.2 were used.

Table 6.2 The basic combination functions from the library used in the presented model

	Notation	**Formula**	**Parameters**

	Notation	Formula	Parameters
Advanced logistic sum	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, \ldots, V_k)$	$\left[\frac{1}{1+e^{-\boldsymbol{\sigma}(V_1+\ldots+V_k-\boldsymbol{\tau})}} - \frac{1}{1+e^{\boldsymbol{\sigma}\boldsymbol{\tau}}}\right](1+e^{-\boldsymbol{\sigma}\boldsymbol{\tau}})$	Steepness $\boldsymbol{\sigma} > 0$ Threshold $\boldsymbol{\sigma}\,\boldsymbol{\tau}^{\boldsymbol{\sigma}}$
Stepmod	$\mathbf{stepmod}_{\boldsymbol{\rho},\boldsymbol{\delta}}(V_1, \ldots, V_k)$	0 if t mod $\boldsymbol{\rho} < \boldsymbol{\delta}$, else 1	Repetition $\boldsymbol{\rho}$ Duration $\boldsymbol{\delta}$

Note that 'network characteristics' and 'network states' are two distinct concepts for a network. Self-modeling is a way to relate these concepts to each other in an interesting and useful way. A *self-model* is making the implicit network characteristics (such as connection weights or excitability thresholds) explicit by adding states to the network representing these characteristics. Thus, the network gets an internal self-model of part of the network structure. When these states are dynamic, this can be used to obtain an *adaptive network*; see (Treur 2020). Such self-models are also very useful to model both simulation of mental models (where the mental model is used) and learning of mental models (where the mental model is changed), as is done in the current chapter. In this way, multiple self-modeling levels can be created where network characteristics from one level relate to states at a next level. This can be used to design *second-order* or *higher-order self-modeling networks*; see, for example, (Treur 2020). More specifically, adding a self-model for a temporal-causal network is done in the way that for some of the states Y of the base network and some of its related network structure characteristics for connectivity, aggregation and timing (i.e., some from $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$), additional network states $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ (self-model states) are introduced:

Connectivity self-model.
Self-model states $\mathbf{W}_{X,Y}$ are added representing connection weights $\boldsymbol{\omega}_{X,Y}$.

Aggregation self-model.
Self-model states $\mathbf{C}_{j,Y}$ are added representing combination function weights $\boldsymbol{\gamma}_{i,Y}$
and/or self-model states $\mathbf{P}_{i,j,Y}$ representing combination function parameters $\boldsymbol{\pi}_{i,j,Y}$

Timing self-model.
Self-model states $\mathbf{H}_Y$ are added representing speed factors $\boldsymbol{\eta}_Y$.

The chosen notations $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ for the self-model states indicate the referencing relation with respect to the characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$: here $\mathbf{W}$ refers to $\boldsymbol{\omega}$, $\mathbf{C}$ refers to $\boldsymbol{\gamma}$, $\mathbf{P}$ refers to $\boldsymbol{\pi}$, and $\mathbf{H}$ refers to $\boldsymbol{\eta}$, respectively. For the processing, these self-model states define the dynamics of state Y in a canonical

manner according to equations in Table 6.2, bottom row, whereby $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$ are replaced by the state values of $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ at time t, respectively.

As the outcome of the addition of a self-model to a temporal-causal network is also a temporal-causal network model itself, as has been shown in (Treur, 2020), Ch 10, this construction can easily be applied iteratively to obtain multiple levels of self-models. Therefore second-order adaptation as, for example, plays an important role to control adaptive processes, can easily be modelled as well. This also has been applied here for the control of the processes in counterfactual thinking.

6.4 A Controlled Adaptive Network Model for Counterfactual Thinking

To explain the introduced network model, the following scenario is used.

Scenario: Jimmy believes he can do an internship & study at the same time (belief). He fails an exam (goal failure), this created an unpleasant situation for him (feeling).

Activation Process: Initially this evokes a mental simulation of the entire event (simulating factual state of affairs) and previous memories are triggered. Activation causes recall of memories in similar situations from past (search space). He focuses on the most relevant memories (nearness); for example: (1) Recalls time he joined private tuitions, (2) Time he studied with friends/study group, (3) Spent extra time on weekends.

Inference Process: Based on mental models for the most relevant memories, Jimmy mentally simulates (nearest) alternatives to the situation and makes evaluations on them.

Learning and Adaptation Process: Once the counterfactuals are inferred, the present situation is re-evaluated by incorporating them in the present situation. This leads to changes in beliefs, as well as action-values (learning). Jimmy's behaviour changes, he joins private tuitions and studies more effectively for next exams (Fig. 6.2).

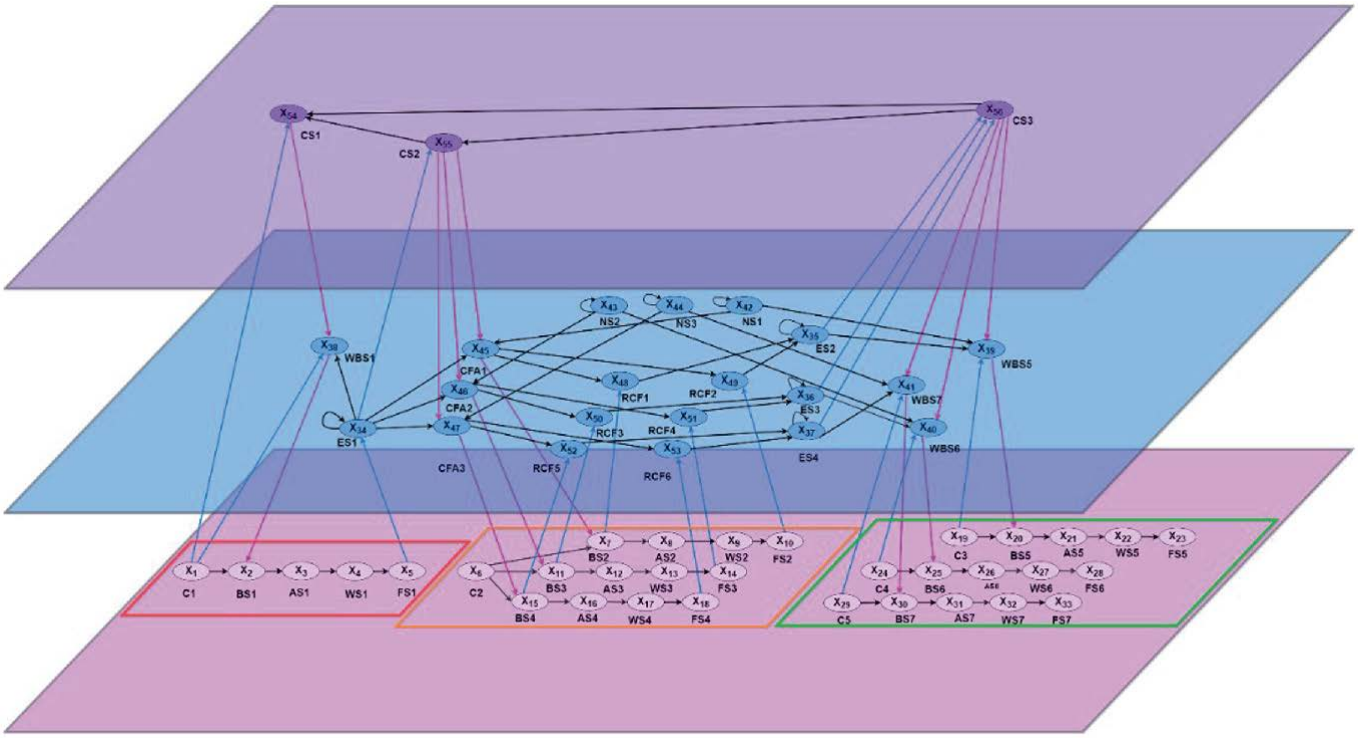


Fig. 6.2 Connectivity of the controlled adaptive network model for counterfactual thinking

The lower base (pink) plane contains the base network where the different colored outlined sections are just added for reasons of presentation to make a clear visual distinction between current states, mental models describing counterfactuals, and future states (C-states, BS-states, AS-states, WS-States, FS-states) involved in the three stages. These colored outlines are not part of the network specification as such. The (blue) plane above the base plane represents the first-order self-model (WBS-states, ES-states, CFA-states, RCF-states, NS-states). The upper (purple) plane above the first-order self-model plane represents the second-order self-model for the control (CS-states). These states at the different levels are explained in more detail in Tables 6.3, 6.4, and 6.5.

Table 6.3 Description of abbreviations used in the model

C	Context States
BS	Belief States
AS	Action States
WS	World States
FS	Feeling States
ES	Evaluation States
WBS	Belief weight representation states for connections from C-states to BS-States
NS	Nearness Indication States
CFA	Counterfactual Activation states
RCF	Representation states for active Counterfactuals
CS	Control States for the three stages Activation, Inference, and Learning and Adaptation

Table 6.4 Overview of the Base States

Base States		Explanation
X_1	C1	Context State 1
X_2	BS1	Belief State 1: 'I can do internship and studies at the same time'
X_3	AS1	Action State 1 for taking up the internship, not studying enough
X_4	WS1	World state 1 for Failing the Exam
X_5	FS1	Feeling state 1: unpleasant
X_6	C2	Context State 2
X_7	BS2	Belief State 2 for Counterfactual thinking: 'Taking private tuitions helps to study well' so 'If I would have taken up private tuitions to prepare, as I had done in the past'
X_8	AS2	Action State 2 for Counterfactual thinking: 'Taking up private tuitions and studying effectively with help from a tutor'
X_9	WS2	World state 2 for BS2: 'Passing the Exam'
X_{10}	FS2	Feeling state 2: 'Feeling pleasant after passing'
X_{11}	BS3	Belief State 3 for Counterfactual thinking: 'one must work extra on the weekends to pass the exam ' so If I would have studied extra on the weekends as well'
X_{12}	AS3	Action State 3 for Counterfactual thinking: 'Studying extra time on weekends as done in the past'
X_{13}	WS3	World state 3 for BS3: 'Passing the Exam'
X_{14}	FS3	Feeling state 3: 'Feeling pleasant after passing'
X_{15}	BS4	Belief State 4 for Counterfactual thinking: 'Studying with friends help to study effectively ' so 'If I would have studied with my friends'
X_{16}	AS4	Action State 4 for Counterfactual thinking: 'studying with friends, learning effectively'
X_{17}	WS4	World state 4 for BS4: 'Passing the Exam'
X_{18}	FS4	Feeling state 4: 'Feeling plesant after passing'
X_{19}	C3	Context State 3 (Future context)
X_{20}	BS5	Belief State 5: 'I can and will have to study harder and study more on the weekends'
X_{21}	AS5	Action State 5 for studying more on the weekends,
X_{22}	WS5	World state 5 for passing the Exam
X_{23}	FS5	Feeling State 5: pleasant
X_{24}	C4	Context State 4 (Future context)
X_{25}	BS6	Belief State 6: 'I can join tuitions, it will help me to study as in the past'
X_{26}	AS6	Action State 6 for joining private tuitions
X_{27}	WS6	World state 6 for Passing the Exam
X_{28}	FS6	Feeling state 6: pleasant

Base States		Explanation
X_1	C1	Context State 1
X_{29}	C5	Context state 5 (Future context)
X_{30}	BS7	Belief state 7: 'I should take help from friends, study with them'
X_{31}	AS7	Action state 7 for taking help from friends
X_{32}	WS7	World state 7 for passing the exam
X_{33}	FS7	Feeling state 7: pleasant

Table 6.5 Overview of the First-Order and Second-Order Self-Model States

Self-Model States		Explanation
X_{34}	ES1	Evaluation state 1 for BS1 & FS1
X_{35}	ES2	Evaluation state 2 for BS2 & FS2 via RCF1 & RCF2
X_{36}	ES3	Evaluation state 3 for BS3 & FS3 via RCF3 & RCF4
X_{37}	ES4	Evaluation state 4 for BS4 & FS4 via RCF5 & RCF6
X_{38}	WBS1	Belief weight representation state for the connection from C1 to BS1
X_{39}	WBS2	Belief weight representation state for the connection from C3 to BS5
X_{40}	WBS3	Belief weight representation state for the connection from C4 to BS6
X_{41}	WBS4	Belief weight representation state for the connection from C5 to BS7
X_{42}	NS1	Nearness Indication State for BS2
X_{43}	NS2	Nearness Indication State for BS3
X_{44}	NS3	Nearness Indication State for BS4
X_{45}	CFA1	Counterfactual Activation State 1 for BS2
X_{46}	CFA2	Counterfactual Activation State 2 for BS3
X_{47}	CFA3	Counterfactual Activation State 3 for BS4
X_{48}	RCF1	Counterfactual Representation State 1 for BS2
X_{49}	RCF2	Counterfactual Representation State 2 for FS2
X_{50}	RCF3	Counterfactual Representation State 3 for BS3
X_{51}	RCF4	Counterfactual Representation State 4 for FS3
X_{52}	RCF5	Counterfactual Representation State 5 for BS4
X_{53}	RCF6	Counterfactual Representation State 6 for FS4
X_{54}	CS1	Control State 1 for stage 1: via activation of WBS1 to BS1
X_{55}	CS2	Control State 2 for stage 2: via activation of CFA2, CFA3, CFA4 to BS2, BS3, BS4
X_{56}	CS3	Control State 3 for stage 3: via activation of WB5, WB6, WB7 to BS5, BS6, BS7

In the network model introduced here, the states about the current situation are represented in the red outlined section of the base plane. As mentioned in the scenario, the current situation leads to unpleasant feeling (FS1) which leads to re-evaluation and then to updating the beliefs. This update of beliefs is modelled by ES-states and WBS-states in the first-order self-model. The NS-states allow to only focus on the mental models for counterfactuals which have small deviations from reality and then choose the best among them (within the orange outlined area in the base plane).

These choices make use of the ES-states ES2 to ES4 for counterfactuals which have links from the active counterfactual representation RCF-states. Based on the (persistent) evaluation states, the learning takes place: the belief weight representation WBS1 is suppressed by evaluation state ES1, and evaluation states ES2 to ES4 make that the belief weight representations WBS5 to WBS7 will become activated as soon as a relevant context occurs. Note that, in addition to the second-order CS-states for control, the first-order CFA-states and WBS-states play a crucial role in control as well. These CFA-states and WBS-states are controlled by the CS-states and in turn they themselves control the related BS-states: BS2 to BS4 by CFA1 to CFA3 and BS5 to BS7 by WBS5 to WBS7. Through this overall control, the processes involved in counterfactual thinking will take place in a structured manner according to the three stages Activation, Inference, and Learning and Adaptation as found from the literature in Sect. 6.2.

6.5 Simulation Results

The computational network model was simulated using a dedicated software environment implemented in MATLAB described in (Treur 2020), Chap. 9. For an example simulation, see Figs. 6.3, 6.4 and 6.5 which all display one and the same simulation but just for the sake of presentation are displayed in parts for the overall processes according to the three stages found in Sect. 6.2. For the simulation $\Delta t = 0.5$ was chosen, the total time 100. The context states are considered external factors and use the **stepmod** function to let them occur at some time. The speed factor for the context states C1 and C2 was set at 0 so that they always are there, whereas for C3 to C5 it was set at 2, and by setting appropriate values for the **stepmod** function's parameters they occur at time 60. For all other states (BS-, AS-, WS-, FS-states) the speed factor was set at 0.5. The connection weights between the states and the other characteristics of the network model and the initial values are shown in Sect. 6.8. All states in the model have initial value 0 except C1, C2 which both have it as 1, and the NS-states (1–3) which have it as specific values depending on the considered variant of the scenario.

In the first stage shown in Fig. 6.3, due to evaluation state ES1, via WBS1 the initial belief state BS1 is suppressed and only BS5 and states that follow it go up representing that an appropriate counterfactual was chosen based on the outcome of the first stage.

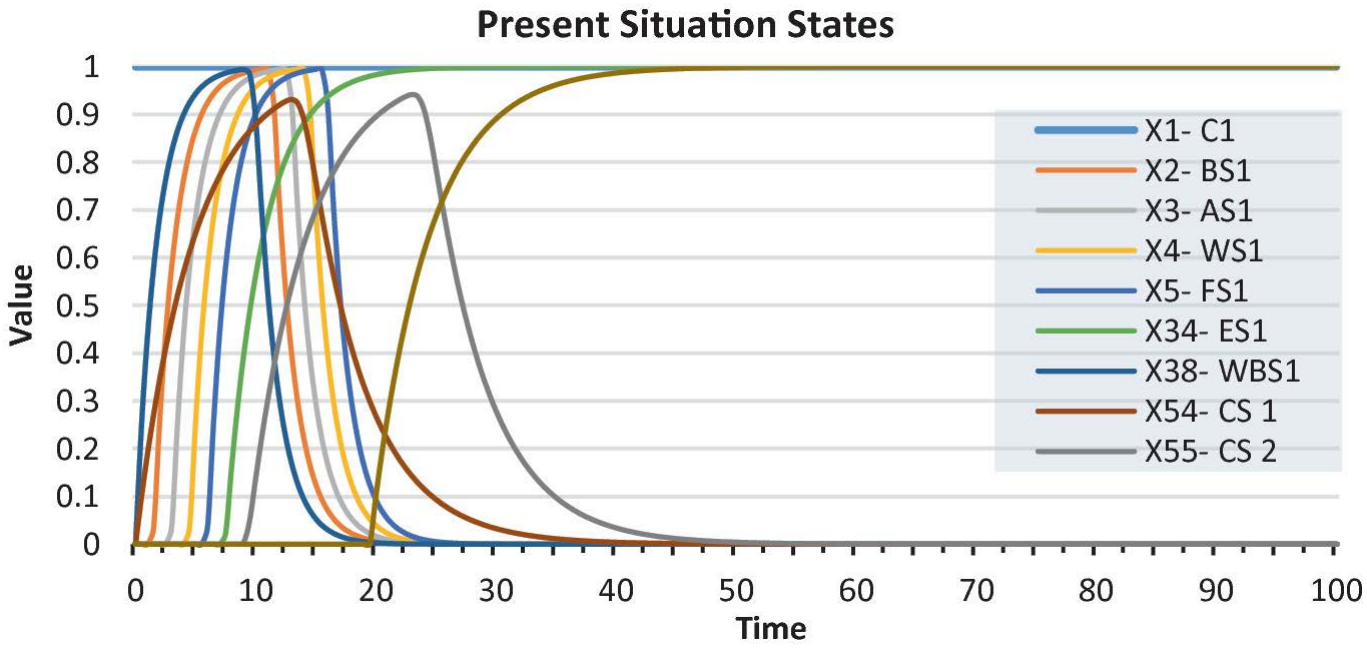


Fig. 6.3 Present Situation States representing the impact of ES-1, WBS-1 and CS-1 on initial belief state BS-1

The second stage is shown in Fig. 6.4. The NS-states are set in such a way that only one of the future states (within the green outline in the base plane) goes for value 1. A typical pattern is that first initial context state and belief state trigger a chain of events leading to activation of ES1. This leads to the activation of different counterfactuals based on their nearness and to their evaluation via evaluation states ES2 to ES4.

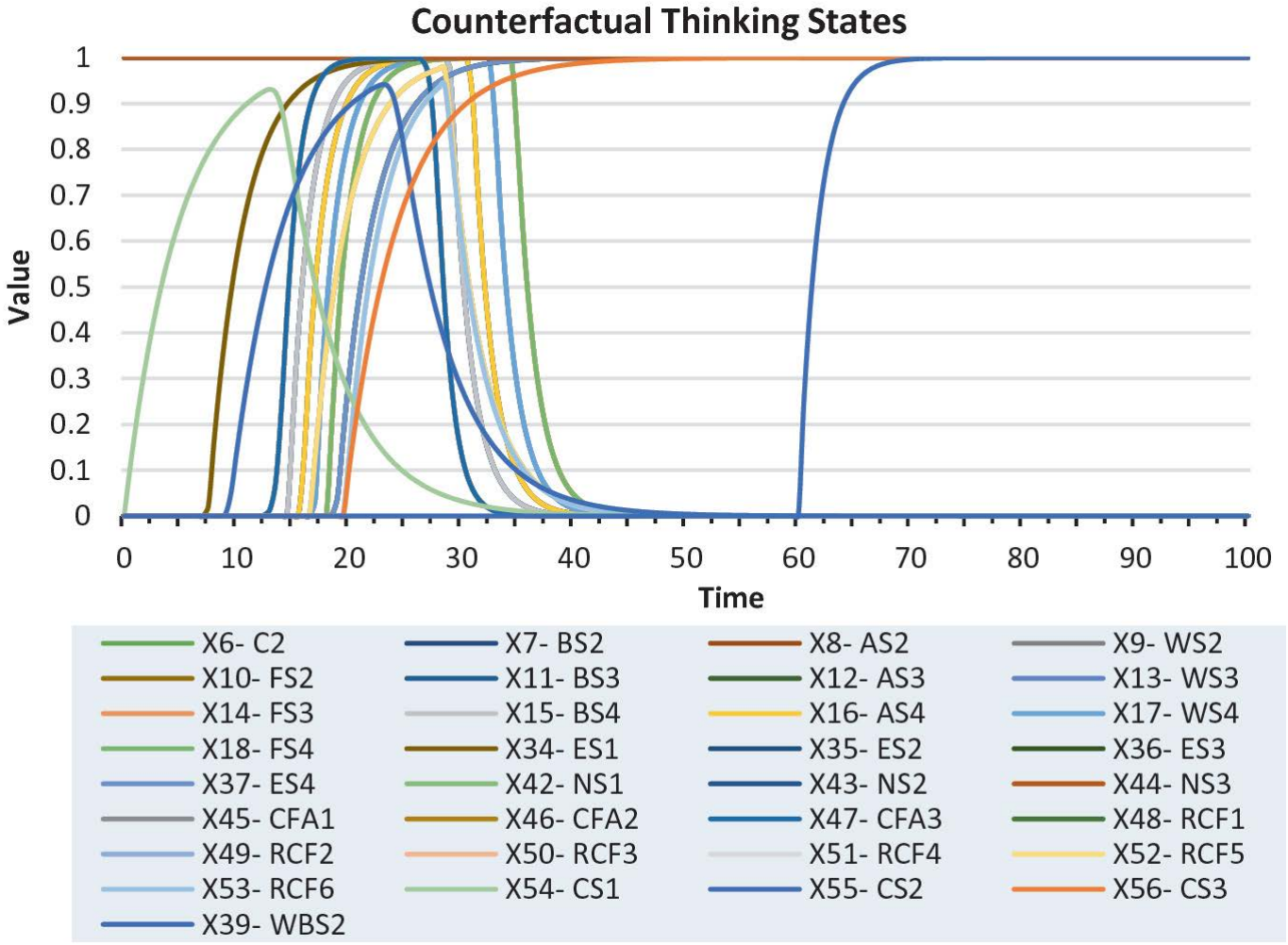


Fig. 6.4 Counterfactual thinking leading to evaluation states, nearness states and updating of beliefs represented by WBS2 going up to 1

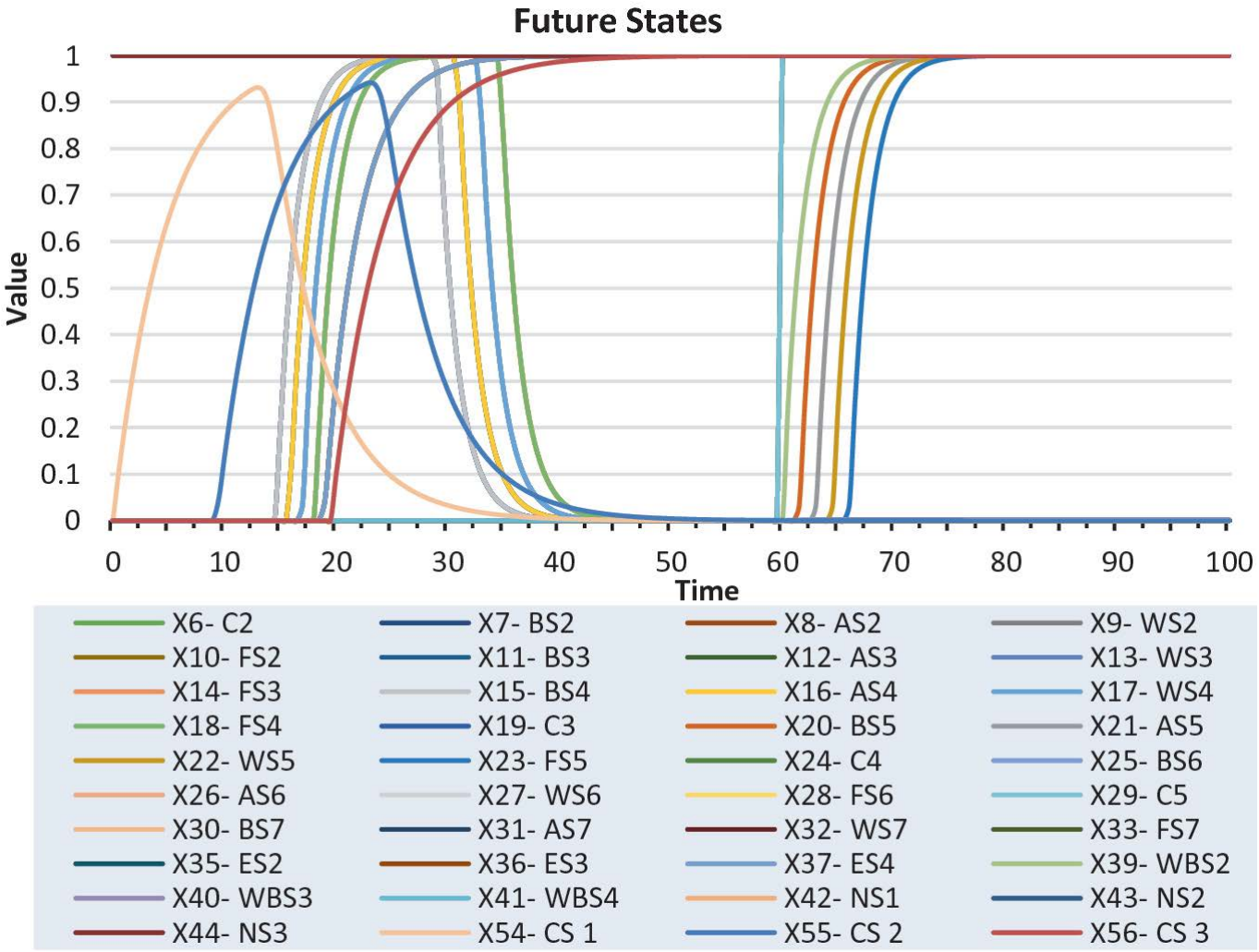


Fig. 6.5 Future states showing the impact from WBS-states and CS-states

mb	base connectivity	1	2	3
X_1	C1	X_1		
X_2	BS1	X_1		
X_3	AS1	X_2		
X_4	WS1	X_3		
X_5	FS1	X_4		
X_6	C2	X_6		
X_7	BS2	X_6		
X_8	AS2	X_7		
X_9	WS2	X_8		
X_{10}	FS2	X_9		
X_{23}	FS5	X_{22}		
X_{24}	C4	X_{24}		
X_{25}	BS6	X_{24}		
X_{26}	AS6	X_{25}		
X_{27}	WS6	X_{26}		
X_{28}	FS6	X_{27}		
X_{29}	C5	X_{29}		
X_{30}	BS7	X_{29}		
X_{31}	AS7	X_{30}		
X_{32}	WS7	X_{31}		
X_{33}	FS7	X_{32}		
X_{34}	ES1	X_5	X_{34}	
X_{35}	ES2	X_{35}	X_{48}	X_{49}
X_{36}	ES3	X_{36}	X_{50}	X_{51}
X_{37}	ES4	X_{37}	X_{52}	
X_{38}	WBS1	X_1	X_{34}	X_{53}
X_{39}	WBS2	X_{19}	X_{35}	X_{42}
X_{11}	BS3	X_6		
X_{12}	AS3	X_{11}		
X_{13}	WS3	X_{12}		
X_{14}	FS3	X_{13}		
X_{15}	BS4	X_6		
X_{16}	AS4	X_{15}		
X_{17}	WS4	X_{16}		
X_{18}	FS4	X_{17}		
X_{19}	C3	X_{19}		
X_{20}	BS5	X_{19}		
X_{21}	AS5	X_{20}		
X_{22}	WS5	X_{21}		
X_{40}	WBS3	X_{24}	X_{36}	X_{43}
X_{41}	WBS4	X_{29}	X_{37}	X_{44}
X_{42}	NS1	X_{42}		
X_{43}	NS2	X_{43}		
X_{44}	NS3	X_{44}		
X_{45}	CFA1	X_{34}	X_{42}	
X_{46}	CFA2	X_{34}	X_{43}	
X_{47}	CFA3	X_{34}	X_{44}	
X_{48}	RCF1	X_7	X_{45}	
X_{49}	RCF2	X_{10}	X_{45}	
X_{50}	RCF3	X_{11}	X_{46}	
X_{51}	RCF4	X_{14}	X_{46}	
X_{52}	RCF5	X_{15}	X_{47}	
X_{53}	RCF6	X_{18}	X_{47}	
X_{54}	CS 1	X_1	X_{55}	X_{56}
X_{55}	CS 2	X_{34}	X_{56}	
X_{56}	CS 3	X_{35}	X_{36}	X_{37}

Fig. 6.6 Connectivity role matrix **mb** for base connectivity

mcw	connection	weights	1	2	3
X_1	C1		1		
X_2	BS1		X_{38}		
X_3	AS1		1		
X_4	WS1		1		
X_5	FS1		1		
X_6	C2		1		
X_7	BS2		X_{45}		
X_8	AS2		1		
X_9	WS2		1		
X_{10}	FS2		1		
X_{11}	BS3		X_{46}		
X_{12}	AS3		1		
X_{13}	WS3		1		
X_{14}	FS3		1		
X_{15}	BS4		X_{47}		
X_{16}	AS4		1		
X_{17}	WS4		1		
X_{18}	FS4		1		
X_{19}	C3		1		
X_{20}	BS5		X_{39}		
X_{21}	AS5		1		
X_{22}	WS5		1		
X_{23}	FS5		1		
X_{24}	C4		1		
X_{25}	BS6		X_{40}		
X_{26}	AS6		1		
X_{27}	WS6		1		
X_{28}	FS6		1		
X_{29}	C5		1		
X_{30}	BS7		X_{41}		
X_{31}	AS7		1		
X_{32}	WS7		1		
X_{33}	FS7		1		
X_{34}	ES1		1	1	
X_{35}	ES2		1	1	1
X_{36}	ES3		1	1	1
X_{37}	ES4		1	1	
X_{38}	WBS1		1	-1	X_{54}
X_{39}	WBS2		1	X_{56}	0.8
X_{40}	WBS3		1	X_{56}	0.3
X_{41}	WBS4		1	X_{56}	0.2
X_{42}	NS1		1		
X_{43}	NS2		1		
X_{44}	NS3		1		
X_{45}	CFA1		X_{45}	1	
X_{46}	CFA2		X_{45}	1	
X_{47}	CFA3		X_{45}	1	
X_{48}	RCF1		1	1	
X_{49}	RCF2		1	1	
X_{50}	RCF3		1	1	
X_{51}	RCF4		1	1	
X_{52}	RCF5		1	1	
X_{53}	RCF6		1	1	
X_{54}	CS 1		1	-1	-1
X_{55}	CS 2		1	-1	
X_{56}	CS 3		1	1	1

Fig. 6.7 Connectivity role matrix **mcw** for connection weights **ω**

For the third stage shown in Fig. 6.5, the evaluation states ES2 to ES4 use their links to the CS3 state which controls the base-level BS5 to BS7 states through WBS5 to WBS7 states to make better future choices than in the past.

6.6 Verification of the Model by Analysis of Stationary Points

To verify whether the implemented network model behaves as expected from its conceptual specification, seven stationary points were analyzed for a simulation example. As a stationary point for a state Y is a point where $dY(t)/dt = 0$, from the equation in Table 6.2, bottom row, the following general criterion for it can be derived: $\boldsymbol{\eta}_Y = 0$ or

$$\mathbf{c}_Y(\boldsymbol{\omega}_{X_1,Y} X_1(t), \ldots, \boldsymbol{\omega}_{X_k,Y} X_k(t)) = Y(t) \tag{6.2}$$

where X_1 to X_k are the states from which Y gets its incoming connections. It has been verified that the aggregated impact $\mathbf{aggimpact}_{X_i}(t)$ defined by the left hand side of (2) matches the state value for some stationary points observed in the example simulation. The results are shown in Table 6.6. The model generated correct values as there were no serious deviations (see the bold numbers) for the stationary points as can be seen from the table: the maximal deviation was 0.001.

Table 6.6 Analysis of stationary points of the model

State X_i	X_{38}-WBS1	X_2-BS1	X_3-AS1	X_4-WS1	X_5-FS1	X_{54}-CF1	X_{52}-RCF5
Time point t	8.48	10.21	11.62	13.02	14.42	12.41	27.29
$X_i(t)$	0.984526	0.98714	0.987113	0.986998	0.986879	0.910897	0.964102
aggimpact$_{Xi}(t)$	0.983823	0.986161	0.98606	0.986017	0.985946	0.909804	0.963555
Deviation	**-0.0007**	**-0.00098**	**-0.00105**	**-0.00098**	**-0.00093**	**-0.00109**	**-0.00055**

6.7 Discussion

In the presented chapter, counterfactual thinking and the internal simulation of mental models involved in it were studied based on empirical literature from Neuroscience and Psychology. Most of the material was adopted from (Bhalwankar and Treur, 2021c). First, a detailed literature review was conducted in identifying processes, neural correlates and theories related to counterfactual thinking from different disciplines. Especially, the dynamic, adaptive and control aspects of the mental model handling in counterfactual thinking were given the attention they deserve, as they are important but often neglected as soon as formalization or computational modeling of counterfactual thinking is addressed. A realistic scenario with respect to counterfactual thinking processes was identified. For example, functional theory of counterfactual thinking views counterfactuals as a beneficial factor in behavior regulation which enhance future performance via different mechanisms (Epstude and Roese 2008; Roese and Epstude 2017). Practically, counterfactuals also aid when they consist of characteristics which focus on better outcomes that alter behavior in a manner consistent with those outcomes.

Based on the literature, a self-modeling temporal-causal network model was designed. This model captures the process of counterfactual thinking and the internal simulation of the mental models involved, including the dynamics, learning and control. Counterfactual thinking has been studied from various perspectives. But, as far as the authors know, a formalized computational model

for it from a neuroscientific perspective including dynamics, adaptation and control of the thoughts and learning was never proposed.

By the implemented adaptive network model, the process of counterfactual thinking was simulated and shown to work as expected from the literature. For the model also some parameter tuning experiments have been performed which produced appropriate parameters as was expected from the literature, and mathematical analysis was conducted that has shown that the implemented model generated correct values compared to the model's design.

The network-oriented modeling approach used here makes it easy to integrate different theories and findings about a phenomenon into a more complex whole. It helps to understand the interactions within the processes. For the current focus, the present study contributes to theory building in understanding the processes of counterfactual thinking from a human-like modeling perspective, thereby taking into account adaptation and control for these processes. This goes much further than what is found in logical approaches to counterfactual reasoning that often abstract most of the dynamic, adaptive and control aspects involved in realistic counterfactual thinking away; e.g., (Starr, 2020). The development of mental models for other types of learning processes has recently been addressed for the case of how mental models for operating a device are formed (Bhalwankar and Treur 2021a, 2021b).

The notion of counterfactual thinking can also be a useful source of inspiration for (not necessarily human-like) AI applications. Counterfactual thinking has already been used as a source of inspiration to refine and optimize well-known Q-learning approaches to reinforcement learning to let agents in a multi-agent setting improve their competitive abilities. A study presented in (Wang et al. 2019), showed that counterfactual thinking can make the agents obtain more accumulative rewards from the environments with fair information in comparison to their opponents.

Further research detailing the selection of the features and evaluation process of counterfactuals can build more accurate human-like or more useful non-human-like models; the process of model building is an iterative one and always ends with an invitation to make a next iteration.

6.8 Appendix: Full Specification of the Adaptive Network Model by Role Matrices.

For an explanation of the concept of role matrices and how they are used as input for the software environment, see (Treur, 2020), Ch. 9 or (Treur, 2022). In Figs. 6.6, 6.7, 6.8, 6.9, 6.10 the full specification of the introduced network model is described.

ms speed factors		1
X_1	C1	0
X_2	BS1	0.5
X_3	AS1	0.5
X_4	WS1	0.5
X_5	FS1	0.5
X_6	C2	0
X_7	BS2	0.5
X_8	AS2	0.5
X_9	WS2	0.5
X_{10}	FS2	0.5
X_{11}	BS3	0.5
X_{12}	AS3	0.5
X_{13}	WS3	0.5
X_{14}	FS3	0.5
X_{15}	BS4	0.5
X_{16}	AS4	0.5
X_{17}	WS4	0.5
X_{18}	FS4	0.5
X_{19}	C3	2
X_{20}	BS5	0.5
X_{21}	AS5	0.5
X_{22}	WS5	0.5
X_{23}	FS5	0.5
X_{24}	C4	2
X_{25}	BS6	0.5
X_{26}	AS6	0.5
X_{27}	WS6	0.5
X_{28}	FS6	0.5
X_{29}	C5	2
X_{30}	BS7	0.5
X_{31}	AS7	0.5
X_{32}	WS7	0.5
X_{33}	FS7	0.5
X_{34}	ES1	0.3
X_{35}	ES2	0.3
X_{36}	ES3	0.3
X_{37}	ES4	0.3
X_{38}	WBS1	0.5
X_{39}	WBS2	0.5
X_{40}	WBS3	0.5
X_{41}	WBS4	0.5
X_{42}	NS1	0
X_{43}	NS2	0
X_{44}	NS3	0
X_{45}	CFA1	0.7
X_{46}	CFA2	0.7
X_{47}	CFA3	0.7
X_{48}	RCF1	0.3
X_{49}	RCF2	0.3
X_{50}	RCF3	0.3
X_{51}	RCF4	0.3
X_{52}	RCF5	0.3
X_{53}	RCF6	0.3
X_{54}	CS 1	0.2
X_{55}	CS 2	0.2
X_{56}	CS 3	0.2

Fig. 6.8 Timing role matrix **ms** for speed factors **η**

mcfw combination weights	combination function	1 stepmod	2 alogistic
X_1	C1	1	
X_2	BS1		1
X_3	AS1		1
X_4	WS1		1
X_5	FS1		1
X_6	C2	1	
X_7	BS2		1
X_8	AS2		1
X_9	WS2		1
X_{10}	FS2		1
X_{11}	BS3		1
X_{12}	AS3		1
X_{13}	WS3		1
X_{14}	FS3		1
X_{15}	BS4		1
X_{16}	AS4		1
X_{17}	WS4		1
X_{39}	WBS2		1
X_{40}	WBS3		1
X_{41}	WBS4		1
X_{42}	NS1		1
X_{43}	NS2		1
X_{44}	NS3		1
X_{45}	CFA1		1
X_{46}	CFA2		1
X_{47}	CFA3		1
X_{48}	RCF1		1
X_{18}	FS4		1
X_{19}	C3	1	
X_{20}	BS5		1
X_{21}	AS5		1
X_{22}	WS5		1
X_{23}	FS5		1
X_{24}	C4	1	
X_{25}	BS6		1
X_{26}	AS6		1
X_{27}	WS6		1
X_{28}	FS6		1
X_{29}	C5	1	
X_{30}	BS7		1
X_{31}	AS7		1
X_{32}	WS7		1
X_{33}	FS7		1
X_{34}	ES1		1
X_{35}	ES2		1
X_{36}	ES3		1
X_{37}	ES4		1
X_{38}	WBS1		1
X_{49}	RCF2		1
X_{50}	RCF3		1
X_{51}	RCF4		1
X_{52}	RCF5		1
X_{53}	RCF6		1
X_{54}	CS 1		1
X_{55}	CS 2		1
X_{56}	CS 3		1

Fig. 6.9 Aggregation role matrix **mcfw** for combination function weights **γ**

mcfpv combination function parameter values		1 stepmod		2 alogistic	
		1	2	1	2
		ρ	δ	σ	τ
X_1	C1	200	0		
X_2	BS1			30	0.5
X_3	AS1			30	0.5
X_4	WS1			30	0.5
X_5	FS1			30	0.5
X_6	C2				
X_7	BS2			30	0.4
X_8	AS2			30	0.4
X_9	WS2			30	0.4
X_{10}	FS2			30	0.4
X_{11}	BS3			30	0.4
X_{12}	AS3			30	0.4
X_{13}	WS3			30	0.4
X_{14}	FS3			30	0.4
X_{15}	BS4			30	0.4
X_{16}	AS4			30	0.4
X_{17}	WS4			30	0.4
X_{18}	FS4			30	0.4
X_{19}	C3	200	60		
X_{20}	BS5			30	0.5
X_{21}	AS5			30	0.5
X_{22}	WS5			30	0.5
X_{23}	FS5			30	0.5
X_{24}	C4	200	60		
X_{25}	BS6			30	0.5
X_{26}	AS6			30	0.5
X_{27}	WS6			30	0.5
X_{28}	FS6			30	0.5
X_{29}	C5	200	60		
X_{30}	BS7			30	0.5
X_{31}	AS7			30	0.5
X_{32}	WS7			30	0.5
X_{33}	FS7			30	0.5
X_{34}	ES1			30	0.5
X_{35}	ES2			30	0.5
X_{36}	ES3			30	0.5
X_{37}	ES4			30	0.5
X_{38}	WBS1			30	0.5
X_{39}	WBS2			30	2.5
X_{40}	WBS3			30	2.5
X_{41}	WBS4			30	2.5
X_{42}	NS1			30	0.5
X_{43}	NS2			30	0.5
X_{44}	NS3			30	0.5
X_{45}	CFA1			30	1.5
X_{46}	CFA2			30	1.5
X_{47}	CFA3			30	1.5
X_{48}	RCF1			30	1.5
X_{49}	RCF2			30	1.5
X_{50}	RCF3			30	1.5
X_{51}	RCF4			30	1.5
X_{52}	RCF5			30	1.5
X_{53}	RCF6			30	1.5
X_{54}	CS 1			30	0.4
X_{55}	CS 2			30	0.4
X_{56}	CS 3			30	0.4

Fig. 6.10 Aggregation role matrix **mcfp** for combination function parameters **ρ**, **δ**, **σ**, **τ**

References

Barbey, A.K., Krueger, F., Grafman, J.: Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. Philosophical Trans. Royal Soc. B: Biol. Sci. **364**(1521), 1291–1300 (2009)
[Crossref]

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, vol. 190(4), pp. 90–101. 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-Pardpo, M. (eds), Proceedings of the 9th International Conference on Complex Networks and Their Applications. Studies in Computational Intelligence, vol. 944, pp. 245–259. Springer Nature Switzerland AG (2021b)

Bhalwankar, R., Treur, J.: 'If only I would have done that...': a controlled adaptive network model for learning by counterfactual thinking. In: Proceedings of the 17th International Conference on Artificial Intelligence Applications and Innovations, AIAI'21, pp. 3–16. Advances in Information and Communication Technology, vol. 627. Springer Nature Switzerland (2021c)

Byrne, R.M.J.: Mental models and counterfactual thoughts about what might have been. Trends Cogn. Sci. **6**(10), 426–431 (2002)
[Crossref]

Byrne, R.M.J.: The Rational Imagination: How People Create Alternatives to Reality. MIT Press (2005)

Byrne, R.M.J.: Precis of 'the rational imagination: how people create alternatives to reality.' Behavior. Brain Sci. **30**(5–6), 439–453 (2007)
[Crossref]

Byrne, R.M.J.: Counterfactual thought. Annu. Rev. Psychol. **67**, 135–157 (2016)
[Crossref]

De Brigard, F., Hanna, E., St Jacques, P.L., Schacter, D.L.: How thinking about what could have been affects how we feel about what was. Cogn. Emot. **33**, 646–659 (2019)
[Crossref]

Epstude, K., Roese, N.J.: The functional theory of counterfactual thinking. Pers. Soc. Psychol. Rev. **12**(2), 168–192 (2008)
[Crossref]

Fortin, N.J., Agster, K.L., Eichenbaum, H.B.: Critical role of the hippocampus in memory for sequences of events. Nat. Neurosci. **5**(5), 458–462 (2002)
[Crossref]

Kahneman, D., Miller, D.T.: Norm theory: comparing reality to its alternatives. Psychol. Rev. **93**(2), 136 (1986)
[Crossref]

Markman, K.D., Gavanski, I., Sherman, S.J., McMullen, M.N.: The mental simulation of better and worse possible worlds. J. Exp. Soc. Psychol. **29**(1), 87–109 (1993)
[Crossref]

Roese, N.J.: The functional basis of counterfactual thinking. J. Pers. Soc. Psychol. **66**(5), 805 (1994)
[Crossref]

Russell, J.A.: Core affect and the psychological construction of emotion. Psychol. Rev. **110**(1), 145 (2003)
[Crossref]

Sanna, L.J., Schwarz, N., Small, E.M.: Accessibility experiences and the hindsight bias: I knew it all along versus it could never have happened. Mem. Cognit. **30**(8), 1288–1296 (2002)
[Crossref]

Starr, W.B.: Conditional and counterfactual logic. In: Knauff, M., Spohn, W. (eds.). The Handbook of Rationality. MIT Press: Cambridge, MA (2020)

Timberlake, B.: The effects of counterfactual comparison on learning and reasoning (Doctoral dissertation, University of Trento) (2019)

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions. Springer Nature, Cham, Switzerland (2016)

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

Treur, J.: With a little help: a modeling environment for self-modeling network models. In: Treur, J., Van Ments, L. (eds.) Mental Models and their Dynamics, Adaptation and Control: a Self-Modeling Network Modeling Approach, Ch. 17 (this volume). Springer Nature (2022)

Tulving, E., Markowitsch, H.J.: Episodic and declarative memory: role of the hippocampus. Hippocampus **8**(3), 198–204 (1998)
[Crossref]

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

Wang, Y., Wan, Y., Zhang, C., Bai, L., Cui, L., Yu, P.: Competitive multi-agent deep reinforcement learning with counterfactual thinking. In 2019 IEEE International Conference on Data Mining (ICDM), pp. 1366–1371. IEEE (2019)

Wilson, R.C., Collins, A.G.: Ten simple rules for the computational modeling of behavioral data. Elife **8**, e49547 (2019)

Roese, N. J., Epstude, K.: The functional theory of counterfactual thinking: New evidence, new challenges, new insights. In Advances in experimental social psychology, vol. 56, pp. 1–79. Academic Press (2017)

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_7

7. Do You Get Me: Controlled Adaptive Mental Models for Analysis and Support Processes

Jan Treur[1] ✉
(1) Social AI Group, Department of Computer Science , Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

✉ **Jan Treur**
Email: j.treur@vu.nl

Abstract
In this chapter, a self-modeling mental network model is presented for cognitive analysis and support processes for a human. These cognitive analysis and support processes are modeled by internal mental models. At the base level, the model is able to perform the analysis and support processes based on these internal mental models. To obtain adaptation of these internal mental models, a first-order self-model is included in the network model. In addition, to obtain control of this adaptation, a second-order self-model is included. This makes the network model a second-order self-modeling network model. The adaptive network model is illustrated for a number of realistic scenarios for a supported car driver.

Keywords Controlled adaptation – Mental model – Analysis and support tasks

7.1 Introduction

To describe complex cognitive processes including their formation, learning, development or adaptation, often the concept of (internal) mental model is used; e.g., (Gentner and Stevens 1983; Greca and Moreira 2000; Kieras and Bovair 1984; Seel 2006). The focus here is on monitoring and assessing the performance of a human in demanding circumstances and generating support actions whenever needed. As an example, when we observed that a driver took alcohol, the assessment will be that it will not be safe to start driving. Based on that, as a

support action, starting the car will be (or tried to be) blocked. Or, if while driving it is observed that the driver's steering is unstable, this will be assessed as a driving risk. Therefore as a support action in this case it will be proposed to slow down the car.

In such analysis and support processes, two internal mental models play a main role: an analysis model to determine assessments based on information obtained from monitoring and a support model to determine proper support actions taking into account generated assessments. In practice, such internal mental models usually are adaptive in order to improve them over time. Moreover, some form of control is applied. The interplay of these three types of processes involving the mental models (applying them, adapting them, and exerting control) forms a complex and adaptive cognitive process. Modeling such a complex adaptive cognitive process is a nontrivial challenge.

A computational model for the considered cognitive processes can be a tool to study how humans perform them, but it can also be a good basis for an AI-application to support a human. Artificial variants of such cognitive processes are currently being built in in new generations of cars as automatic safety systems.

This chapter addresses how the complex cognitive processes considered here can be modeled using second-order self-modeling networks (Treur 2020a, b). The three elements mentioned above (applying the mental models, adapting them, and exerting control) are addressed by three levels in such a self-modeling network. At the base level within the self-modeling network model, the internal mental models are modeled and executed, at the first self-modeling level the adaptation of these mental models is modeled, and at the second self-modeling level exerted control is modeled.

In the chapter, in Sect. 7.2 self-modeling networks are briefly introduced. Section 7.3 addresses the application domain. Section 7.4 presents the design of the considered self-modeling network model and Sect. 7.5 presents example simulation scenarios of it. Section 7.6 is a discussion.

7.2 Network Models Using Self-models

In this section, the network-oriented modeling approach used from (Treur 2020a, b) is briefly summarised.

Distinction between network characteristics and network states

The following is a crucial distinction for network models:

- Network *characteristics* (such as connection weights and excitability thresholds) have values (their strengths) and determine (e.g., cognitive)

processes and behaviour in an implicit, automatic manner. They can be considered to provide an *embodiment view* on the network. In principle, these characteristics by themselves may not be directly accessible nor observable for network states (or a person: usually you do not see or feel a specific connection in your brain).

- Network *states* (such as sensor states, sensory representation states, preparation states, emotion states) have values (their activation levels) and are explicit representations that may be accessible for network states or a person and can be handled or manipulated explicitly. They can be considered to provide an *informational view* on the network; usually the states are assumed to have a certain informational content. In principle, for the case of a mental network, states may be accessible or observable for a person: you may see (mental image), feel (emotion) or note in some other way a specific state in your brain.

Following (Treur 2016, 2020b), a temporal-causal network model is characterised by (here X and Y denote nodes of the network, also called states):

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

The following difference (or differential) equations that are used for simulation purposes and also for analysis of temporal-causal networks incorporate these network characteristics $\boldsymbol{\omega}_{X,Y}$, $\mathbf{c}_Y(..)$, $\boldsymbol{\eta}_Y$ in a standard numerical format:

$$Y(t+\Delta t) = Y(t) + \boldsymbol{\eta}_Y \left[\mathbf{c}_Y\left(\boldsymbol{\omega}_{X_1,Y}X_1(t) \quad , ..., \boldsymbol{\omega}_{X_k,Y}X_k(t)\right) - Y(t)\right]\Delta t \quad (7.1)$$

for any state Y and where X_1 to X_k are the states from which Y gets its incoming connections. Here the overall combination function $\mathbf{c}_Y(..)$ for state Y is the weighted average of available basic combination functions $\mathbf{c}_j(..)$ by specified weights $\boldsymbol{\gamma}_{j,Y}$ (and parameters $\boldsymbol{\pi}_{1,j,Y}$, $\boldsymbol{\pi}_{1,j,Y}$ of $\mathbf{c}_j(..)$) for Y:

(7.2)

$$\mathbf{c}_Y\left(V_1,\ldots,V_k\right)=\frac{\boldsymbol{\gamma}_{1,Y}\mathbf{c}_1\left(V_1,\ldots,V_k\right)+\cdots+\boldsymbol{\gamma}_{m,Y}\mathbf{c}_m\left(V_1,\ldots,V_k\right)}{\boldsymbol{\gamma}_{1,Y}+\cdots+\boldsymbol{\gamma}_{m,Y}}$$

Such Eqs. (7.1), (7.2) and the ones in Table 7.1 are hidden in the dedicated software environment; see (Treur 2020b), Ch 9. Within the software environment described there, a large number of around 40 useful basic combination functions are included in a combination function library; see Table 7.1 for the first two of them: these are the ones used in this chapter. The above concepts enable to design network models and their dynamics in a declarative manner, based on mathematically defined functions and relations. How it works is that the network characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{j,Y}$, $\boldsymbol{\pi}_{1,j,Y}$, $\boldsymbol{\pi}_{1,j,Y}$, $\boldsymbol{\eta}_Y$ that define the design of the network model, are given as input to the dedicated software environment, and hidden within this environment the difference Eqs. (7.1) are executed for all states, thus generating simulation graphs as output.

Table. 7.1 Basic combination functions from the library used in the model presented here

	Notation	**Formula**	**Parameters**
Euclidean	$\mathbf{eucl}_{\boldsymbol{n},\boldsymbol{\lambda}}(V_1, \ldots, V_k)$	$\sqrt[n]{\frac{V_1^n+\cdots+V_k^n}{\lambda}}$	Order $\boldsymbol{n} > 0$ Scaling factor $\boldsymbol{\lambda} > 0$
Advanced logistic sum	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, \ldots, V_k)$	$\left[\frac{1}{1+\mathrm{e}^{-\boldsymbol{\sigma}(V_1+\cdots+V_k-\boldsymbol{\tau})}}-\frac{1}{1+\mathrm{e}^{\boldsymbol{\sigma}\boldsymbol{\tau}}}\right](1+\mathrm{e}^{-\boldsymbol{\sigma}\boldsymbol{\tau}})$	Steepness $\boldsymbol{\sigma} > 0$ Excitability threshold $\boldsymbol{\tau}$

Self-models representing network characteristics by network states

The self-modeling network modeling approach is inspired by the more general idea of self-referencing or 'Mise en abyme', sometimes also called 'the Droste-effect' after the famous Dutch chocolate brand who uses this effect in packaging and advertising of their products already since 1904. For some examples, see Fig. 7.1. For more explanation, see for example, https://en.wikipedia.org/wiki/Mise_en_abyme, https://en.wikipedia.org/wiki/Droste_effect. This effect occurs in art when within artwork a small copy of the same artwork is included. This can be applied graphically in paintings or photographs, or in sculptures. Also, it is sometimes used within literature (story-within-the-story), theater (theater-within-the-theater), or movies (movie-within-the-movie).

Fig. 7.1 Three examples of the Mise en abyme or Droste-effect. http://michel.parpere.pagesperso-orange.fr/pedago/voc/mise%20en%20abyme.htm, https://www.instagram.com/culturfemale/, https://www.instagram.com/p/CCYmVLMpGPo/

This idea is applied to network models as follows. As indicated above, 'network characteristics' and 'network states' are two distinct concepts for a network. Self-modeling is a way to relate these distinct concepts to each other in an interesting and useful way:

- A *self-model* is making the implicit network characteristics (such as connection weights and excitability thresholds) explicit by adding states for these characteristics; thus the network gets an internal self-model of part of the network structure of itself.
- In this way, different self-modeling levels can be created where network characteristics from one level relate to explicit network states at a next level. By iteration, an arbitrary number of self-modeling levels can be modeled, covering *second-order or higher-order effects*.

Self-modeling networks can be recognized both in physical and mental domains. For example:

- In the *physical domain*, in the brain, information about the characteristics of the network of causal relations between activation states of neurons is, for example, represented in physical configurations for synapses (e.g., connection weights), neurons (e.g., excitability thresholds) and/or chemical substances (e.g., neurotransmitters).
- In the *mental domain*, a person can create mental states in the form of representations of his or her own (personal) characteristics, thus forming a subjective self-model (acquired by experiences); e.g., of being very sensitive for pain or for critical feedback or of having an anger issue.

Adding a self-model for a temporal-causal network is done in the way that for some of the states Y of the base network and some of the network structure characteristics for connectivity, aggregation and timing (in particular, some from $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$), additional network states $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ (self-model states) are introduced (see the blue upper plane in Fig. 7.2 and further):

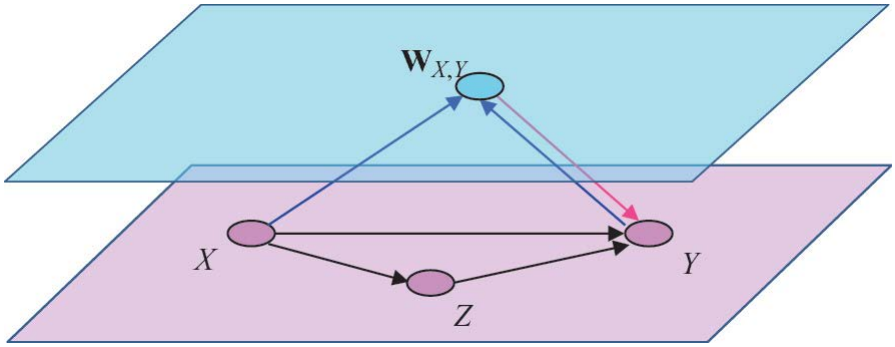


Fig. 7.2 Connectivity characteristics of the self-model for the Hebbian Learning adaptation principle

(a)

Connectivity self-model

- Self-model states $\mathbf{W}_{X_i,Y}$ are added representing connectivity characteristics, in particular connection weights $\boldsymbol{\omega}_{X_i,Y}$

(b)

Aggregation self-model

- Self-model states $\mathbf{C}_{j,Y}$ are added representing aggregation characteristics, in particular combination function weights $\boldsymbol{\gamma}_{i,Y}$
- Self-model states $\mathbf{P}_{i,j,Y}$ are added representing aggregation characteristics, in particular combination function parameters $\boldsymbol{\pi}_{i,j,Y}$

(c)

Timing self-model

- Self-model states $\mathbf{H}_Y$ are added representing timing characteristics, in particular speed factors $\boldsymbol{\eta}_Y$.

The notations $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ for the self-model states indicate the referencing relation with respect to the characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$: here $\mathbf{W}$ refers to $\boldsymbol{\omega}$, $\mathbf{C}$ refers to $\boldsymbol{\gamma}$, $\mathbf{P}$ refers to $\boldsymbol{\pi}$, and $\mathbf{H}$ refers to $\boldsymbol{\eta}$, respectively. For the processing, these self-model states define the dynamics of state Y in a canonical manner according to Eq. (7.1) whereby $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$ are replaced by the state values of $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ at time t, respectively.

An example of an aggregation self-model state $\mathbf{P}_{i,j,Y}$ for a combination function parameter $\boldsymbol{\pi}_{i,j,Y}$ is for the excitability threshold $\boldsymbol{\tau}_Y$ of state Y, which is the second parameter of the logistic sum combination function; then $\mathbf{P}_{i,j,Y}$ is usually indicated by $\mathbf{T}_Y$, which refers to threshold $\boldsymbol{\tau}_Y$. Such aggregation self-model states $\mathbf{T}_Y$ will play an important role in the network model addressed below, as will connectivity self-model states $\mathbf{W}_{X,Y}$, referring to connection weights $\boldsymbol{\omega}_{X,Y}$. Similarly, self-model states $\mathbf{H}_Y$ can be added that refer to the speed factor $\boldsymbol{\eta}_Y$ of Y.

As the outcome of the addition of a self-model is also a temporal-causal network model itself, as has been proven in (Treur 2020b), Ch 10, this construction can easily be applied iteratively to obtain multiple orders of self-models. This is applied by adding second-order self-model states $\mathbf{H}_{\mathbf{WT}Y}$ representing the adaptive speed factors (i.e., adaptive learning rates in this case) for all first-order self-model states $\mathbf{T}_Y$ and $\mathbf{W}_{X,Y}$ which in turn represent the adaptive threshold $\boldsymbol{\tau}_Y$ of Y and the adaptative connection weights $\boldsymbol{\omega}_{X,Y}$ of all incoming connections of Y.

7.3 Modeling the Adaptation Principles Used

In this section, it will be shown how the modeling approach for self-modeling network models described in Sect. 7.2 has been applied to model the adaptation principles of first- and second order used here. When self-models are changing over time in a proper manner, this offers a useful method to model any adaptation principle. This does not only apply to first-order adaptive networks, but also to second-order adaptive networks, modeling control by using second-order self-models.

7.3.1 First-Order Self-models for the First-Order Adaptation Principles Used

Within Cognitive Neuroscience literature, two types of first-order adaptation are considered, one for connection weights and one for intrinsic neuronal properties; for example, as described in (Chandra and Barkai 2018):

> Learning-related cellular changes can be divided into two general groups: modifications that occur at synapses and modifications in the intrinsic properties of the neurons. While it is commonly agreed that changes in strength of connections between neurons in the relevant networks underlie memory storage, ample evidence suggests that modifications in intrinsic neuronal properties may also account for learning related behavioral changes. Chandra and Barkai (2018, p. 30)

In this chapter for these two types of adaptivity, two first-order adaptation principles are considered: Hebbian Learning for connection weights and Excitability Modulation for the excitability threshold of states.

The Hebbian Learning adaptation principle

A well-known adaptation principle of the first type (addressing adaptive connectivity) is Hebbian Learning (Hebb 1949), which can be explained by:

> When an axon of cell A is near enough to excite B and repeatedly or persistently (7.3)
>
> takes part in firing it, some growth process or metabolic change takes place in one
>
> or both cells such that A's efficiency, as one of the cells firing B, is increased. (Hebb 1949, p. 62)

This is sometimes simplified (neglecting the phrase 'one of the cells firing B') to:

> What fires together, wires together. (Shatz 1992; Keysers and Gazzola 2014) (7.4)

Within a self-modeling network, this can be modeled by using a *connectivity self-model* based on self-model states $\mathbf{W}_{X,Y}$ representing connection weights $\boldsymbol{\omega}_{X,Y}$. These self-model states need incoming and outgoing connection to let them function within the network. To incorporate the 'firing together' part, for the self-model's connectivity, incoming connections from X and Y to $\mathbf{W}_{X,Y}$ are used; see Fig. 7.2 (upward arrows in blue). These upward connections have weight 1 here. Also a connection from $\mathbf{W}_{X,Y}$ to itself with weight 1 is used to model persistence of the learnt effect; in pictures they are usually left out. In addition, an outgoing connection from $\mathbf{W}_{X,Y}$ to state Y is used to indicate where this self-model state $\mathbf{W}_{X,Y}$ has its effect; see Fig. 7.2 (pink downward arrow). The downward connection indicates that the value of $\mathbf{W}_{X,Y}$ is actually used for the connection weight of the connection from X to Y. For the *aggregation characteristics* of the self-model, one of the options for a learning rule is defined by the combination function $\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W)$ from Table 7.2; note that $\mathbf{hebbneg}_{\boldsymbol{\mu}}(V_1, V_2, W)$ is a similar variant of Hebbian Learning for connections with negative weights. For more options of Hebbian Learning combination functions and further mathematical analysis of their limit behaviour, see, for example (Treur 2020b), Ch. 14.

Table. 7.2 Combination functions for self-models modeling first- and second-order adaptation principles used here

Name and self-model state	Combination functions	Variables and parameters
Hebbian Learning $\mathbf{W}_{X,Y}$	$\mathbf{hebb}_{\boldsymbol{\mu}}(V_1, V_2, W) = V_1 V_2 (1 - W) + \mu W$ $\mathbf{hebbneg}_{\boldsymbol{\mu}}(V_1, V_2, W) = -V_1(1 - V_2)(1 + W) + \mu W$	V_1,V_2 activation levels of connected states W activation level of self-model state $\mathbf{W}_{X,Y}$ for connection weight $\boldsymbol{\omega}$ $\boldsymbol{\mu}$ persistence factor
Excitability Modulation $\mathbf{T}_Y$	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, ..., V_k)$	$V_1, ..., V_k$ impacts from base states
Exposure Accelerates Adaptation $\mathbf{H}_{\mathbf{W}\mathbf{T}_Y}$	$\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(V_1, ..., V_k)$	$V_1, ..., V_k$ impacts from base states and first-order self-model states

The Excitability Modulation adaptation principle

Although connectivity adaptation is often addressed in the literature, also other characteristics can be made adaptive such as excitability thresholds. For example, the following quote indicates that synaptic activity relates to long-lasting modifications in excitability of neurons:

> Long-lasting modifications in intrinsic excitability are manifested in changes (7.5)
> in the neuron's response to a given extrinsic current (generated by synaptic
> activity or applied via the recording electrode). (Chandra and Barkai 2018, p. 30)

For more literature on this form of learning or adaptation (called here the Excitability Modulation adaptation principle), see (Aizenman and Linden 2000; Daoudal and Debanne 2003; Debanne et al. 2019; Lisman et al. 2018; Titley et al. 2017; Zhang and Linden 2003). Since the adaptation depends on activation of the base states of a state Y and the states X, Z from which it gets its incoming connections, this can be modeled in a self-modeling network in a similar form as above using a self-model state $\mathbf{T}_Y$, as shown in Fig. 7.3.

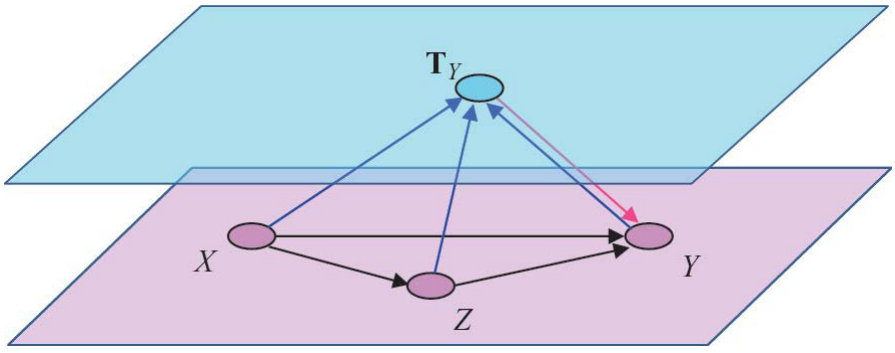


Fig. 7.3 Connectivity characteristics of a self-model for the Excitability Modulation adaptation principle

In this case, based on literature such as (Aizenman and Linden 2000; Chandra and Barkai 2018; Daoudal and Debanne 2003; Debanne et al. 2019; Lisman et al. 2018; Titley et al. 2017; Zhang and Linden 2003) it is assumed that exposure enhances excitability, which means that it decreases the excitability threshold. To achieve this, for the self-model state $\mathbf{T}_Y$ a monotonically increasing combination function can be used, while the connection weights from X, Y, Z to $\mathbf{T}_Y$ are negative; examples of monotonically increasing combination functions are the logistic sum functions and the Euclidean function (with odd order $\boldsymbol{n}$) from Table 7.1. In this case, the (pink) downward connection from $\mathbf{T}_Y$ to Y indicates that the value of $\mathbf{T}_Y$ is used for the threshold value of the logistic sum function of base state Y.

7.3.2 Second-Order Self-model for the Second-Order Adaptation Principle

The first-order adaptation principles discussed in Sect. 7.3.1 refer to forms of *plasticity*. It was shown how they can be described by a first-order self-model for connectivity or aggregation characteristics of the base network, in particular for the connection weights and/or the excitability thresholds used in aggregation. Under which circumstances and to which extent such plasticity actually takes place is controlled by a form of so-called *metaplasticity*; e.g., (Abraham and Bear 1996; Garcia 2002; Magerl et al. 2018; Robinson et al. 2016; Sehgal et al. 2013; Sjöström et al. 2008). Such control can address 'The Plasticity Versus Stability Conundrum' (e.g., Sjöström et al. 2008, p. 773) by only making plasticity happen in circumstances when it is important for the person to change and otherwise stabilize it. Here we consider the following specific second-order adaptation principle for such control of first-order adaptation.

The Adaptation Accelerates with Increasing Exposure adaptation principle

For example, in (Robinson et al. 2016) the following compact quote is found indicating that increasing stimulus exposure makes that the adaptation speed increases:

> Adaptation accelerates with increasing stimulus exposure (7.6).
> (Robinson et al. 2016, p. 2)

This indeed describes a form of metaplasticity that controls the speed of adaptation (learning rate). This principle can be modeled by a (dynamic) second-order self-model for timing characteristics (speed factors) of a first-order self-model for the first-order adaptation. Such a second-order is based on self-model

states $\mathbf{H}_{\mathbf{W}_{X,Y}}$ or $\mathbf{H}_{\mathbf{T}_Y}$ for adaptive learning speed of any of the two types of learning discussed in Sect. 7.3.1, or $\mathbf{H}_{\mathbf{WT}_Y}$ for both types combined. The principle formulated by (6) indicates that the activation level of these second-order self-model states should depend in a monotonically increasing manner on the activation levels of the base states involved: these base states are Y itself and the states *X*, *Z* from which *Y* gets an incoming connections. This makes that the connectivity of this timing self-model (for both forms of learning) is as shown in Fig. 7.4: the (positive, blue) upward connections from the base states *X, Y* and *Z* to the self-model state $\mathbf{H}_{\mathbf{WT}_Y}$ are used to express the part of the principle in (6) referring to 'stimulus exposure'. For the aggregation, for $\mathbf{H}_{\mathbf{WT}_Y}$, an Euclidean (with odd order ***n***) or a logistic sum function can be used to get the monotonic effect as needed. The (negative, blue) upward connections from $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ to the self-model state $\mathbf{H}_{\mathbf{WT}_{X,Y}}$ indicates a counterbalancing effect that makes that the learning speed is limited depending on the learnt level as represented by $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$. The downward (pink) connections from $\mathbf{H}_{\mathbf{WT}_Y}$ to $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ indicate that the value of $\mathbf{H}_{\mathbf{WT}_Y}$ is actually used as speed factor for $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$.

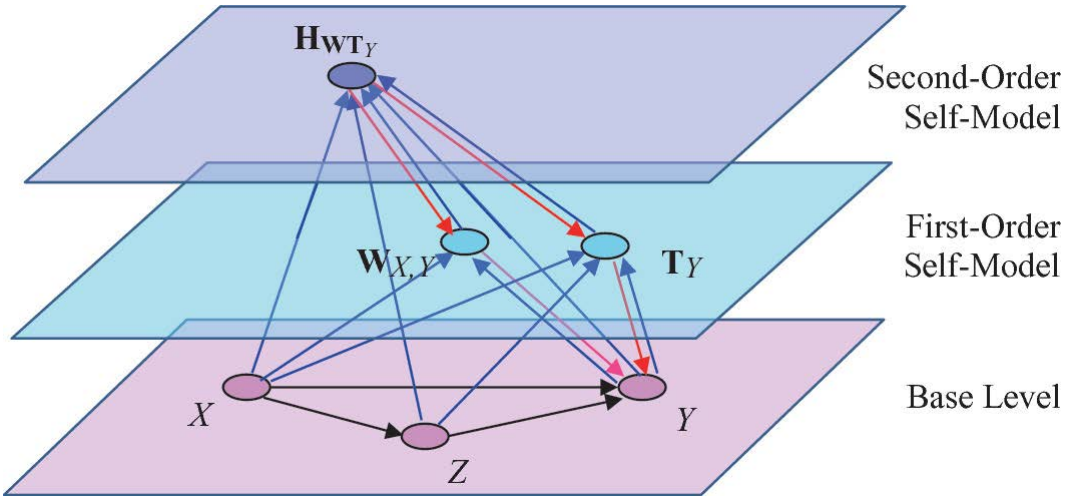


Fig. 7.4 Connectivity of a second-order self-model for the second-order Exposure Accelerates Adaptation adaptation principle for control of first-order self-models for Hebbian Learning and Excitability Modulation

This shows how a specific self-modeling network model is obtained according to a more general three-level self-modeling network design for handling internal mental models with the following three levels:

- Base level:
  applying the mental models (the pink plane in Fig. 7.4)
- First-order adaptation level:
  adapting the mental models (the blue plane in Fig. 7.4)

- Second-order adaptation level:
  exerting control over the adaptation of the mental models (the purple plane in Fig. 7.4)

7.4 Analysis and Support Processes

In situations where humans perform complex demanding tasks, it may be better to keep an eye on them, to monitor how they are performing and to assess their performance. If performance gets poor, support actions may be considered. The mental processes to determine such assessments and to determine appropriate support actions when needed are complex cognitive processes. In the car driver example considered here, it is assumed that continuously sensor or observation data are available. This may concern information about the driver's alcohol usage, his or her gaze and steering behaviour and the amount of rest taken. An unfocused gaze or unstable steering behaviour may be assessed as a driving risk. If that assessment occurs, a support action may be needed, such as slowing down the car or advice to do that. The knowledge behind these mental processes may be adaptive, enabling that the underlying processes will improve over time. Within such complex adaptive cognitive processes, usually internal mental models are used; e.g., (Gentner and Stevens 1983; Greca and Moreira 2000; Kieras and Bovair 1984; Seel 2006). In the case addressed here such internal mental models address (see also Fig. 7.5):

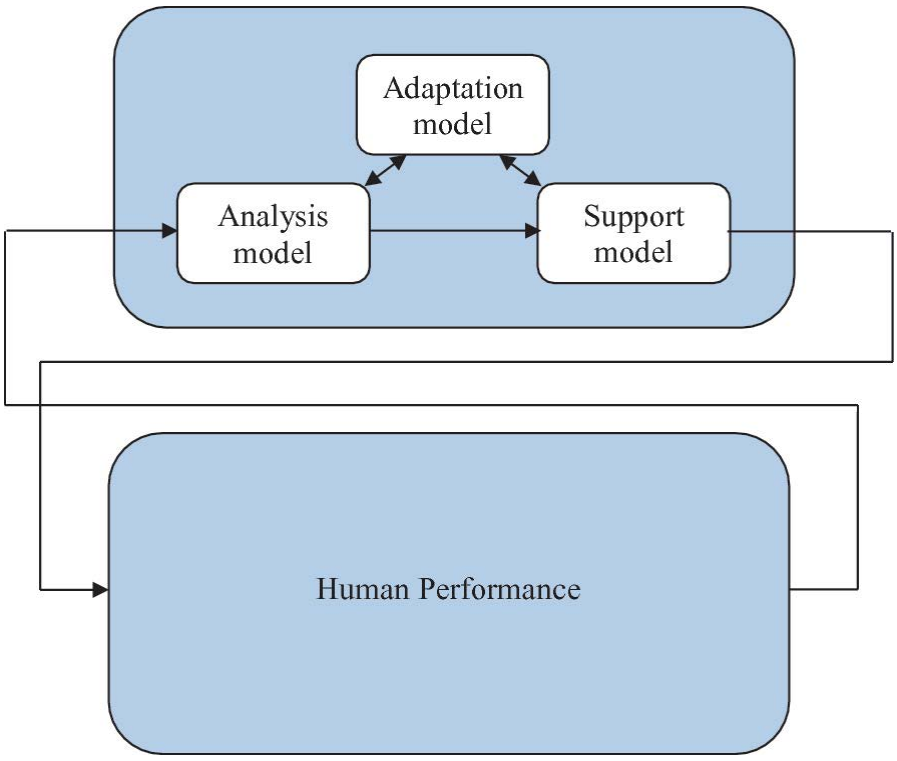


Fig. 7.5 Adaptive model-based architecture to analyse and support humans; adapted from (Treur 2016), Ch 16, p. 469

- *analysis model*
  This model is used for the assessment process or task of the human's performance using observations (e.g., using specific sensors) and domain knowledge. Examples used in the car driver example are a long period of driving, a not well-focused gaze, unstable steering, and alcohol usage. Examples of generated assessments are a risk for getting exhausted or (other) driving risks.
- *support model*
  This model is used for the process or task to generate support actions based on the assessments and domain knowledge. Examples of support actions that are generated by this process are advice to take a rest period, blocking the starting of the car (when not driving), and slowing down the car (when driving).

  As such processes are in principle adaptive to enable improvement of them, a third internal mental model is needed (Treur 2016), Ch. 16.
- *adaptation model*
  To get the analysis and support model better fitting to the specific characteristics of the situation including the driver and car. This works via adapting certain characteristics of the internal mental models.

Section 7.5 addresses how these internal mental models and the way they are used can be modeled by a self-modeling network, leading to the second-order self-modeling network model that is proposed.

7.5 The Second-Order Adaptive Network Model

In this section it is described how the modeling approach discussed in Sect. 7.2 has been used to model the adaptive mental models for analysis and support of human performance from Sect. 7.3 within a self-modeling network.

A useful network architecture to handle internal mental models in general is a self-modeling network that covers at least two levels; see also (Bhalwankar and Treur 2021):

- a base level representing the mental model as a network so that it can be applied or executed (based on the mental model's *within-network dynamics*)
- a first-order self-model explicitly representing the (network) characteristics of the mental model which can be used for formation and adaptation of the mental model (adding *dynamics of* the mental model).

As discussed in Sect. 7.3.2, in human processes the extent to which plasticity actually occurs is controlled by a form of metaplasticity; e.g., (Abraham and Bear 1996; Garcia 2002; Magerl et al. 2018; Robinson et al. 2016; Sehgal et al. 2013; Sjöström et al. 2008). Therefore, in addition a third level (see Sect. 7.3.2) is included with.

- a second-order self-model to control these adaptation processes (control of network adaptation)

This provides a formal computational model of the general three-level network architecture shown in Fig. 7.6, that is applied here; see also (Van Ments and Treur 2021). A specific example of this was already shown in Fig. 7.4. Note that by the upward interlevel connections (the blue upward arrows), this general network architecture enables the use of context-specific information from the mental model at the base level for the learning and context-specific information from both lower levels for the control. This allows to arrange that both the learning and the control take place in a context-sensitive manner.

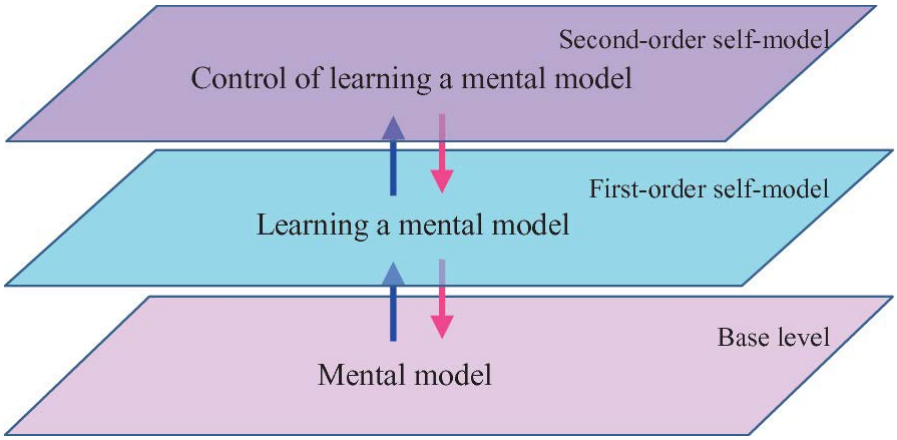


Fig. 7.6 The general three-level network architecture applied

For reasons of presentation, the introduced model will be discussed in two steps. First, Fig. 7.7 displays the connectivity of the first two levels of the introduced network model: the mental models at the base level within the base (pink) plane, and the first-order self-model within the upper (blue) plane. Table 7.3 provides an overview of all states; here the states X_1 to X_{10} model the base level and states X_{11} to X_{25} the network's first-order self-model. The second step concerns the second-order self-model level; this will be described in Section 7.5.3.

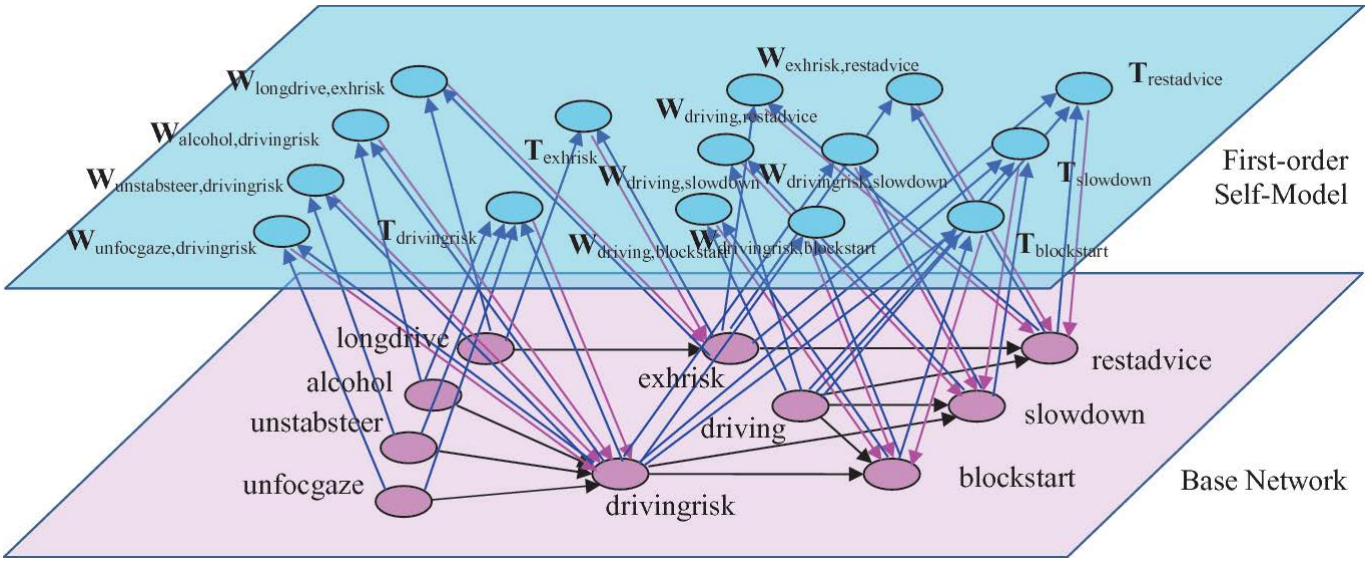


Fig. 7.7 Connectivity for the first two levels of the self-modeling network model

Table 7.3 Explanation of the states of the second-order self-modeling network model

Name	Explanation
X_1 longdrive	The driver is driving for a long period of time
X_2 alcohol	Alcohol is detected
X_3 unstabsteer	The driver's steering is unstable
X_4 unfocgaze	The driver's gaze is not focused
X_5 exhrisk	Assessment of a risk that the driver will get exhausted
X_6 drivingrisk	Assessment of a safety risk for the driving
X_7 driving	The car is driving
X_8 restadvice	Supporting action to advice the driver to take some rest
X_9 slowdown	Supporting action to slow down the car
X_{10} blockstart	Supporting action to block the starting of the car
X_{11} $\mathbf{W}_{\text{longdrive,exhrisk}}$	First-order connectivity self-model state for weight of the connection from longdrive to exhrisk
X_{12} $\mathbf{W}_{\text{alcohol,drivingrisk}}$	First-order connectivity self-model state for weight of the connection from alcohol to drivingrisk
X_{13} $\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	First-order connectivity self-model state for weight of the connection from unstablesteer to drivingrisk
X_{14} $\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	First-order connectivity self-model state for weight of the connection from longdrive to drivingrisk
X_{15} $\mathbf{T}_{\text{exhrisk}}$	First-order aggregation self-model state for excitability threshold of exhrisk
X_{16} $\mathbf{T}_{\text{drivingrisk}}$	First-order aggregation self-model state for excitability threshold of drivingrisk
X_{17} $\mathbf{W}_{\text{exhrisk,restadvice}}$	First-order connectivity self-model state for weight of the connection from exhrisk to restadvice
X_{18} $\mathbf{W}_{\text{driving,restadvice}}$	First-order connectivity self-model state for weight of the connection from driving to restadvice
X_{19} $\mathbf{W}_{\text{drivingrisk,slowdown}}$	First-order connectivity self-model state for weight of the connection from drivingrisk to slowdown
X_{20} $\mathbf{W}_{\text{driving,slowdown}}$	First-order connectivity self-model state for weight of the connection from driving to slowdown
X_{21} $\mathbf{W}_{\text{drivingrisk,blockstart}}$	First-order connectivity self-model state for weight of the connection from drivingrisk to blockstart
X_{22} $\mathbf{W}_{\text{driving,blockstart}}$	First-order connectivity self-model state for weight of the connection from driving to blockstart
X_{23} $\mathbf{T}_{\text{restadvice}}$	First-order aggregation self-model state for excitability threshold of restadvice
X_{24} $\mathbf{T}_{\text{slowdown}}$	First-order aggregation self-model state for excitability threshold of slowdown
X_{25} $\mathbf{T}_{\text{blockstart}}$	First-order aggregation self-model state for excitability threshold of blockstart
X_{26} $\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$	Second-order connectivity self-model state for weight of the connection from longdrive to $\mathbf{T}_{\text{exhrisk}}$
X_{27} $\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$	Second-order connectivity self-model state for weight of the connection from exhrisk to $\mathbf{T}_{\text{exhrisk}}$
X_{28} $\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$	Second-order connectivity self-model state for weight of the connection from alcohol to $\mathbf{T}_{\text{drivingrisk}}$
X_{29} $\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$	Second-order connectivity self-model state for weight of the connection from unstabsteer to $\mathbf{T}_{\text{drivingrisk}}$
X_{30} $\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$	Second-order connectivity self-model state for weight of the connection from unfocgaze to $\mathbf{T}_{\text{drivingrisk}}$
X_{31} $\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$	Second-order connectivity self-model state for weight of the connection from drivingrisk to $\mathbf{T}_{\text{drivingrisk}}$
X_{32} $\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$	Second-order connectivity self-model state for weight of the connection from exhrisk to $\mathbf{T}_{\text{restadvice}}$
X_{33} $\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$	Second-order connectivity self-model state for weight of the connection from restadvice to $\mathbf{T}_{\text{restadvice}}$
X_{34} $\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$	Second-order connectivity self-model state for weight of the connection from drivingrisk to $\mathbf{T}_{\text{slowdown}}$
X_{35} $\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$	Second-order connectivity self-model state for weight of the connection from driving to $\mathbf{T}_{\text{slowdown}}$
X_{36} $\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$	Second-order connectivity self-model state for weight of the connection from slowdown to $\mathbf{T}_{\text{slowdown}}$
X_{37} $\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$	Second-order connectivity self-model state for weight of the connection from drivingrisk to $\mathbf{T}_{\text{blockstart}}$
X_{38} $\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$	Second-order connectivity self-model state for weight of the connection from driving to $\mathbf{T}_{\text{blockstart}}$
X_{39} $\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$	Second-order connectivity self-model state for weight of the connection from blockstart to $\mathbf{T}_{\text{blockstart}}$
X_{40} $\mathbf{HWT}_{\text{exhrisk}}$	Second-order timing self-model state for the speed of states $\mathbf{W}_{X,\text{exhrisk}}$ and $\mathbf{T}_{\text{exhrisk}}$
X_{41} $\mathbf{HWT}_{\text{drivingrisk}}$	Second-order timing self-model state for the speed of states $\mathbf{W}_{X,\text{drivingrisk}}$ and $\mathbf{T}_{\text{drivingrisk}}$
X_{42} $\mathbf{HWT}_{\text{restadvice}}$	Second-order timing self-model state for the speed of states $\mathbf{W}_{X,\text{restadvice}}$ and $\mathbf{T}_{\text{restadvice}}$
X_{43} $\mathbf{HWT}_{\text{slowdown}}$	Second-order timing self-model state for the speed of states $\mathbf{W}_{X,\text{slowdown}}$ and $\mathbf{T}_{\text{slowdown}}$
X_{44} $\mathbf{HWT}_{\text{blockstart}}$	Second-order timing self-model state for the speed of states $\mathbf{W}_{X,\text{blockstart}}$ and $\mathbf{T}_{\text{blockstart}}$

7.5.1 The Base Level

At the base level (see lower, pink plane in Fig. 7.7), in the first place a number of context representation states are included: the states for a long time of driving, alcohol usage, unstable steering, and unfocused gaze (X_1 to X_4, respectively), and for driving (state X_7) in contrast to standing still. These states represent the

situation that is considered and are assumed to be available through sensoring or observation. In addition, two states (X_5 and X_6) are available for assessments exhaustion risk and driving risk, and three states (X_8 to X_{10}) for the support actions rest advice, slow down, and block start. These assessments have incoming connections from the context representation states (X_1 to X_4) on which they depend and the support actions have incoming connections from the assessments and from context representation state X_7 for driving. All these connections have adaptive weights $\boldsymbol{\omega}_{X,Y}$. The assessment states and support action states use the combination function $\mathbf{alogistic}_{\boldsymbol{\sigma},\boldsymbol{\tau}}(..)$ of which the excitability threshold $\boldsymbol{\tau}$ is adaptive.

7.5.2 First-Order Self-models

Within the base network two subnetworks can be distinguished, one for a mental model for analysis to determine an assessment of the performance and one for a mental model to determine support actions. For assessment, for the sake of simplicity the considered scenarios include the following two options: exhaustiveness risk and driving risk. For this, input information is used on long drive, alcohol, unstable steering, and unfocused gaze. The mental model for the support process uses also as input the assessments and generates support actions for which the three options are: rest advice, slow down, and block start. For the two mental models at the base level, self-models have been added which enables adaptation or learning of them (in the upper plane in Fig. 7.7):

First-Order Self-Model for the Analysis Process
First-order self-model **W**-states and **T**-states X_{11} to X_{16} in Table 7.3.

First-Order Self-Model for the Support Process
First-order self-model **W**-states and **T**-states X_{17} to X_{25} in Table 7.3

These self-models represent the relevant network characteristics for connectivity (**W**-states for connection weights) and for aggregation (**T**-states for excitability thresholds) of the two mental model networks at the base level.

Each of these first-order self-model states $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ has a downward connection (in pink) to indicate the state Y of the mental model at the base level for which they have their special effect; so, in relation to these downward links, the value of $\mathbf{W}_{X,Y}$ plays the role of the indicated connection weight and the value of $\mathbf{T}_Y$ the role of the indicated excitability threshold. Each of these first-order self-model states $\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ has upward incoming connections to give it the relevant information about activation of the base level states, as the adaptation

depends on that information via the first-order adaptation principles for Hebbian Learning (3), (4) and for Excitability Modulation (5) discussed in Sect. 7.3. This enables these self-model states to be dynamic according to the indicated adaptation principles, thereby using the appropriate combination functions as indicated in Table 7.2.

Note that a negative weight is used for the connections to $\mathbf{T}_Y$ from the states within the base level that causally precede the indicated state Y. This makes that $\mathbf{T}_Y$ gets lower values when values of these base level states are higher. To counterbalance this negative effect on $\mathbf{T}_Y$, a positive weight is used for the upward connection from Y itself to $\mathbf{T}_Y$. These two opposite effects create a context-sensitive equilibrium value for each aggregation self-model state $\mathbf{T}_Y$. In this way, each aggregation self-model based on **T**-states learns and adapts to the context.

7.5.3 Second-Order Self-models

To incorporate control of the adaptation, a self-model has been added for the first-order self-models that model the adaptation, as shown (in the purple plane) in Fig. 7.8. This makes that the learning (in particular the learning speed) is adaptive itself. As self-model of a self-model, this is a second-order self-model. Figure 7.8 displays the connectivity of the complete second-order adaptive network model. The second-order self-model states in this upper (purple) plane are the states exerting control over adaptation:

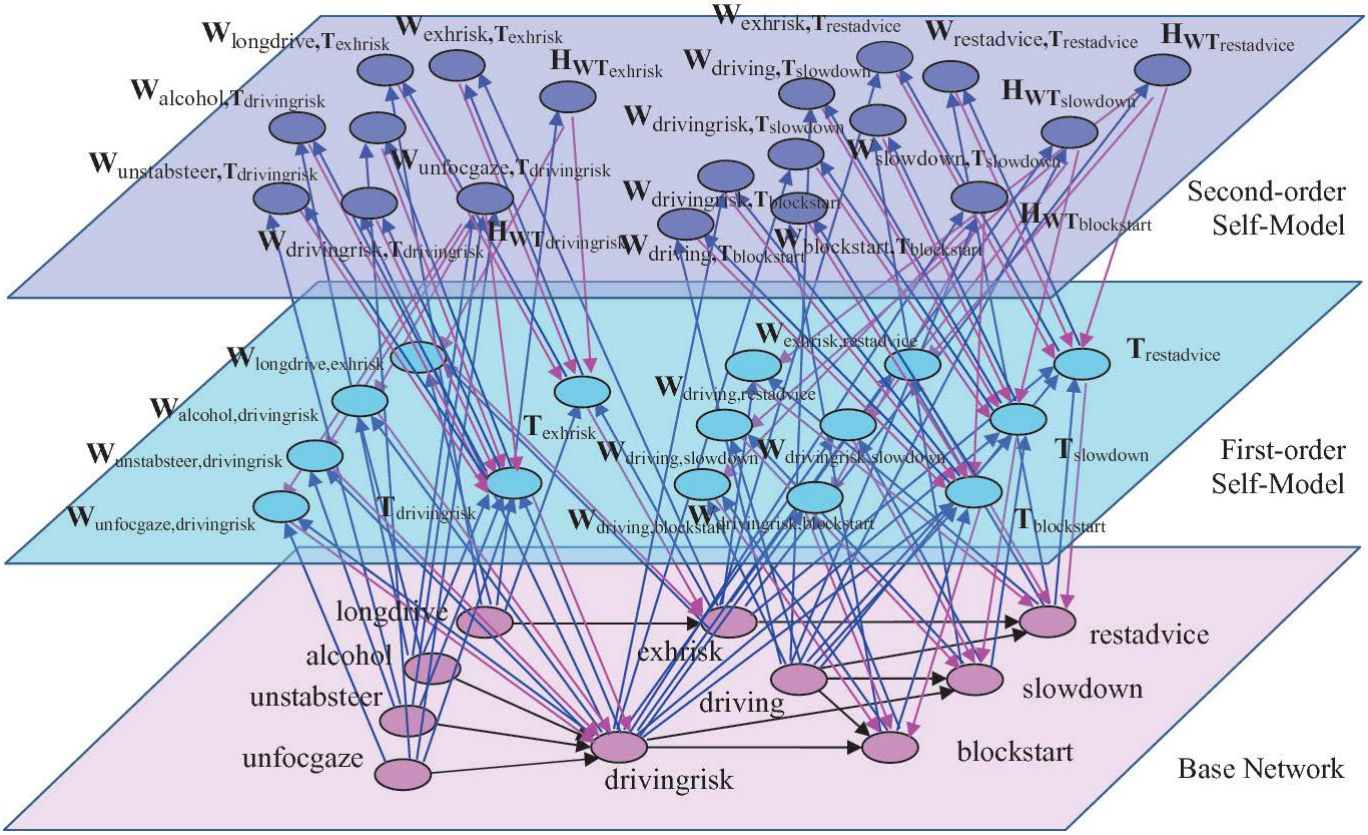


Fig. 7.8 Connectivity of the complete second-order self-modeling network model

Second-Order Self-Models for the Adaptation Process

Second-order self-model **W**-states and $\mathbf{H_{WT}}$-states X_{26} to X_{44} in Table 7.3

The second-order self-model consists of two parts:

- a second-order *connectivity self-model* using states $\mathbf{W}_{X,T_Y}$ representing the weights of the incoming connections of the **T**-states of the first-order self-model
- a second-order *timing self-model* using states $\mathbf{H}_{\mathbf{WT}_Y}$ representing the speed factors of the first-order **W**-states and **T**-states.

These second-order self-models are dynamic, which makes the whole network a second-order adaptive network. To achieve this, the states $\mathbf{H}_{\mathbf{WT}_Y}$ are affected by upward connections from the base level network, following the second-order adaptation principle Exposure Accelerates Adaptation (6) for metaplasticity; e.g., (Robinson et al. 2016; Sjöström et al. 2008). To this end, there are (blue) upward links to each state $\mathbf{H}_{\mathbf{WT}_Y}$ from the base states causally preceding base state *Y*. As these connections get positive weights, when these causal 'antecedents' of *Y* get higher activation levels, the adaptation speed will increase as well. The special effect of each state $\mathbf{H}_{\mathbf{WT}_Y}$ as speed factor for states

$\mathbf{W}_{X,Y}$ and $\mathbf{T}_Y$ is indicated by the downward (pink) connection to the related state $\mathbf{T}_Y$.

The states $\mathbf{W}_{X,T_Y}$ in the second-order self-model change according to Hebbian Learning (3), (4), similar to the states $\mathbf{W}_{X,Y}$ in the first-order self-model as described in Sect. 7.5.2. The second-order self-model states $\mathbf{H}_{\mathbf{W}\mathbf{T}_Y}$ and $\mathbf{W}_{X,T_Y}$ together exert control over the adaptation process modeled by the first-order self-model. Hereby, the former type of states control the speed of adaptation (learning rate) and the latter type of states control which thresholds of the mental models are adapted and how much. All this control takes place in a context-sensitive manner, as via their incoming connections these second-order self-model states are affected by context-specific information from the lower levels.

7.6 Simulation Scenarios

In this section the simulation results for a number of realistic scenarios are discussed. Two different types of scenarios are discussed in Sects. 7.6.1 and 7.6.2, respectively:

- Only adaptive excitability thresholds, connection weights not adaptive
- Both adaptive excitability thresholds and adaptive connection weights

7.6.1 Using Adaptive Excitability Thresholds and Constant Connection Weights

In Figs. 7.9, 7.10 and 7.11 simulation results are shown for three realistic scenarios where the excitability threshold self-model states are dynamic but the connection weight self-model states are not: all connection weight self-model states have constant values assigned. They are defined by the common settings as shown in the role matrices in the Appendix in Sect. 7.8 (the first variant) and specific constant values 0 or 1 for the context representation states X_1 to X_4 and X_7 as shown in Table 7.4. These graphs display the following.

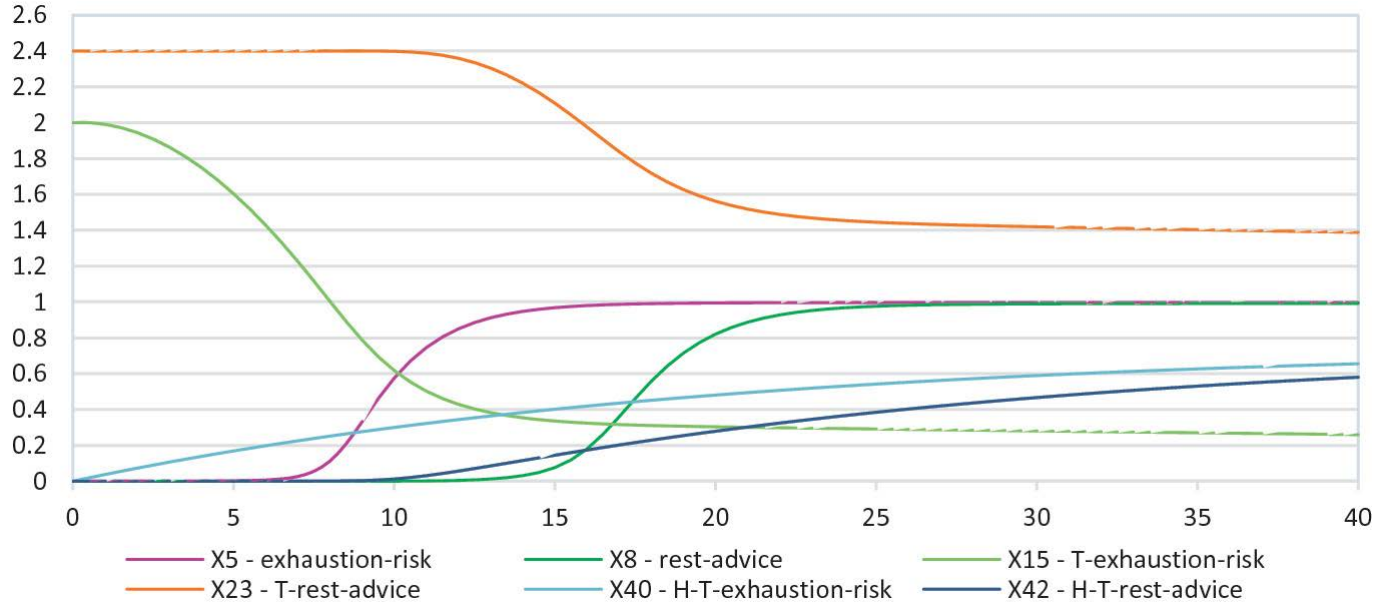


Fig. 7.9 Long drive leads to an exhaustion risk assessment and to the support action rest advice

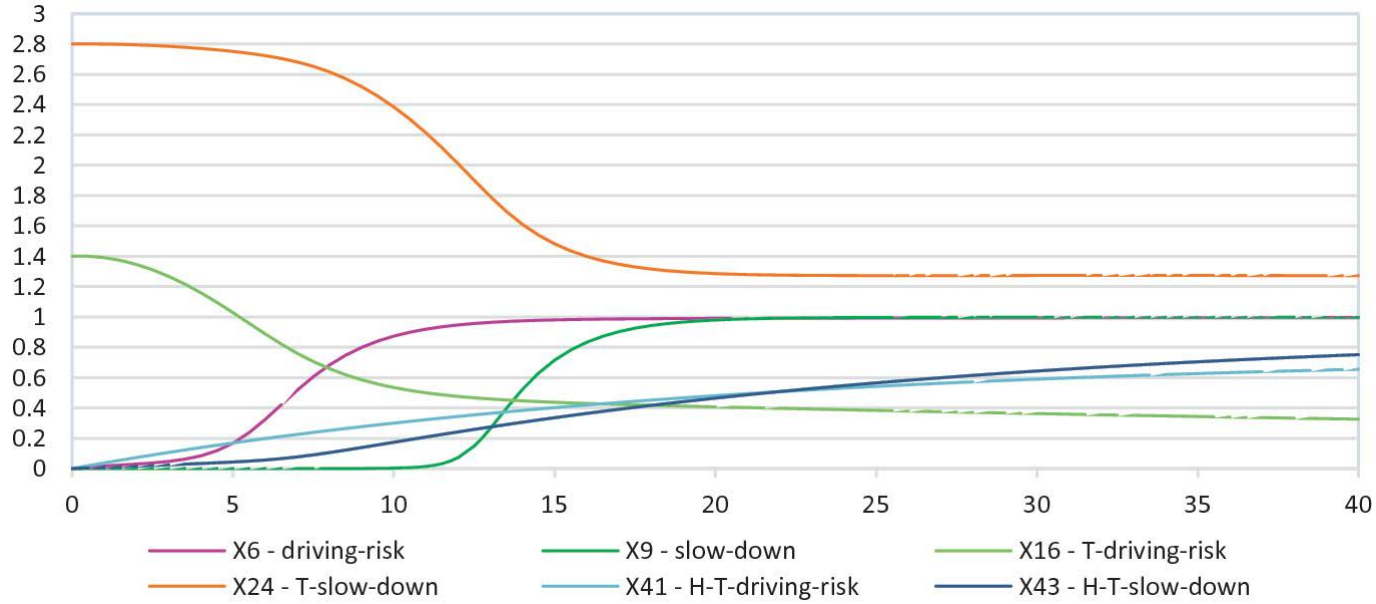


Fig. 7.10 Driving with an unfocused gaze leads to a driving risk assessment and to the support action slow down

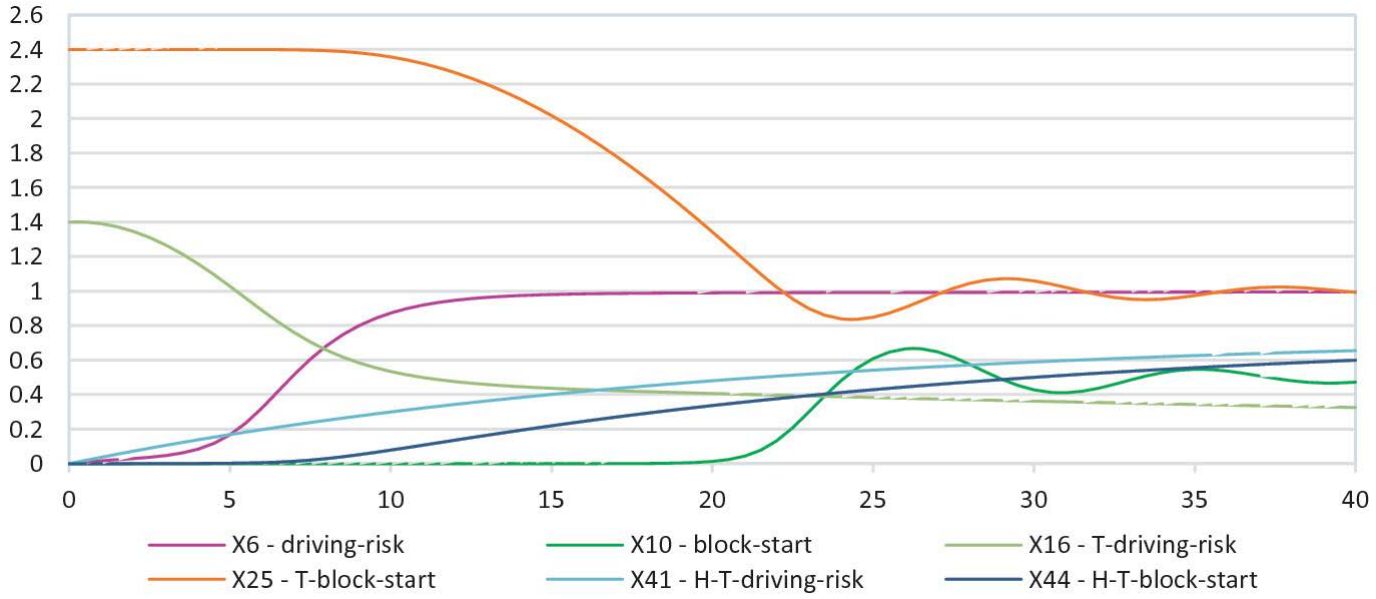


Fig. 7.11 Alcohol usage leads to a driving risk assessment and to the support action block start

Table. 7.4 The three displayed scenarios

Scenario		**Scenario 1.1 Fig. 7.9**	**Scenario 1.2 Fig. 7.10**	**Scenario 1.3 Fig. 7.11**
Explanation		**Driving for a too long time**	**Driving with an unfocused gaze**	**Blocked driving after consuming alcohol**
X_1	longdrive	1	0	0
X_2	alcohol	0	0	1
X_3	unstabsteer	0	0	0
X_4	unfocgaze	0	1	0
X_7	driving	1	1	0

- **For the base level**:
  how the the assessment is generated by the analysis model and how the support action is generated by the support model
- **For the first-order adaptation level**:
  how the excitability thresholds used within the analysis model and the support model adapt over time.
- **For the second-order adaptation or control level**:
  how the adaptation speeds for the adaptations change over time.

Initially the values for the excitability thresholds were set high, in order to illustrate the adaptation process that was needed to get good results. Moreover, the adaptation speed values were initially set at 0. Therefore, in the first phase nothing happens at the base level until the adaptation speeds increase, causing the adaptation process to start. In a next phase this results in successful

adaptation of the analysis and support models, after which they can generate the appropriate outcomes at the base level.

Scenario 1.1: Driving too long

In Fig. 7.9 for Scenario 1.1, it is shown that by the second-order self-model the adaptation speed for the exhaustion risk excitability threshold increases from time 0 on (see the purple line). This is because the long driving input present from time 0 on works as a stimulus. That the adaptation speed increases with this stimulus is in accordance with the second-order adaptation principle Exposure Accelerates Adaptation [14] discussed in Sect. 7.3.2. Moreover, this is also conforming to 'The Plasticity Versus Stability Conundrum' discussed in (Sjöström et al. 2008), p. 773: only adapt (adaptation speed >0) when relevant, otherwise keep stable (adaptation speed 0). The observed increase in adaptation speed results in actual adaptation of this excitability threshold, conform to (5) from Sect. 7.3.1; see (Chandra and Barkai 2018): starting at value 2, it decreases to (after time 13) get values between 0.2 and 0.4 (the brown line). It is clear that this is low enough: after time 10 the exhaustion risk assessment is generated and reaches value 1 after time 15 (the pink line). This makes that a successful analysis model outcome is achieved.

In turn this outcome makes that after time 10 the adaptation model increases the adaptation speed for the excitability threshold of the support action *rest advice* in the support model (the orange line). This results in actual adaptation of that threshold: the value (initially was 2.4) decreases after time 10 and reaches values between 1.4 and 1.6 after time 18 (the dark purple line). This is low enough, as the support action rest advice comes up after time 18 and reaches 1 after time 25 (the dark green line). This makes that a successful support model outcome is achieved.

Scenario 1.2: Driving with an unfocused gaze

In Fig. 7.10 for Scenario 1.2, it is shown that by the adaptation model the adaptation speed for the driving risk excitability threshold (within the analysis model) increases from time 0 on (the light blue line). This leads to adaptation of this threshold: starting at value 1.4, it decreases to (after time 7) reach values below 0.7 (the light green line).

Due to this, from time 5–10 the driving risk assessment is generated and reaches value 1 after time 15 (the pink line). This makes that a successful analysis model outcome is achieved. In turn this makes that by the adaptation model after time 5 the adaptation speed for the excitability threshold of the support action *slowdown* in the support model gets higher (the dark green line). This results in

actual adaptation of that threshold: the value which initially was set at 2.8 decreases after time 10 and reaches values between 1.4 and 1.6 after time 18 (middle green line). Due to this, the support action slow down increases after time 18 and reaches 1 after time 20 (the brown line). This makes that a successful support model outcome is achieved.

Scenario 1.3: Blocking start after alcohol usage

For Scenario 1.3, Fig. 7.11 shows a similar initial pattern as in Scenario 1.2. However, later on (for the support model), this scenario shows that also a fluctuating pattern can occur. This illustrates how the adaptation of the excitability threshold gets reinforcement from the outcome of the support model, so that in the end they reach an equilibrium according to a fluctuating pattern.

7.6.2 Using Both Adaptive Excitability Thresholds and Connection Weights

Next, a number of simulations for example scenarios are discussed where not only the excitability thresholds are adaptive but also the connection weights. The role matrices for these scenarios can be found in the appendix in Sect. 7.8 (the second variant). The values for the context states can be found in Table 7.5.

Table. 7.5 The three displayed scenarios for double adaptivity

Scenario		**Scenario 2.1 Fig. 7.12**	**Scenario 2.2 Fig. 7.13**	**Scenario 2.3 Figs. 7.14, and 7.15**	
Explanation		**Driving too long**	**Driving with unstable steering**	**Driving and alcohol**	
				Nondriving	**Driving**
X_1	longdrive	1	0	0	0
X_2	alcohol	0	0	1	1
X_3	unstabsteer	0	0	0	0
X_4	unfocgaze	0	1	0	0
X_7	driving	1	1	0	1

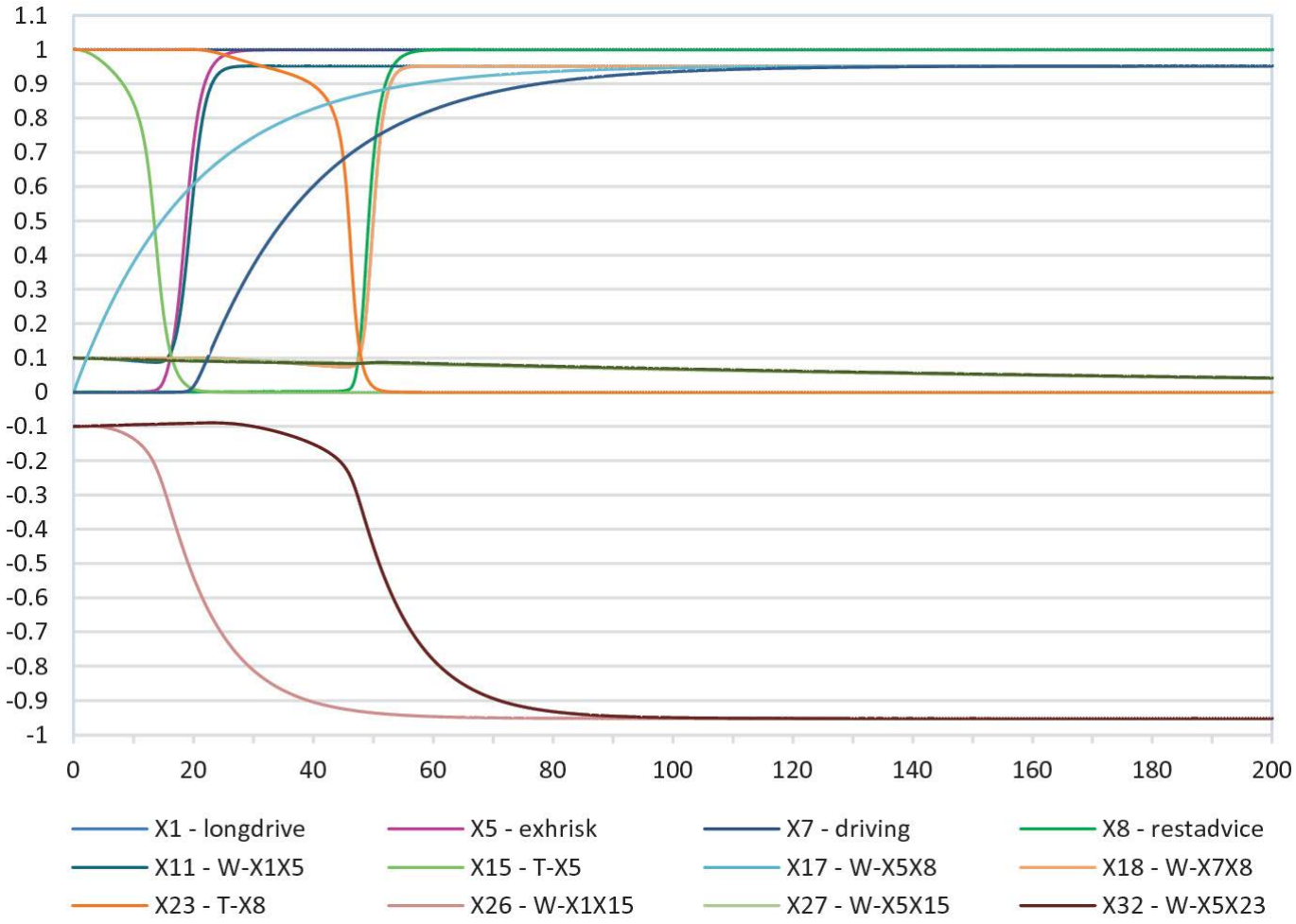


Fig. 7.12 Too long driving leads to a rest advice support action with both adaptive excitability and connectivity

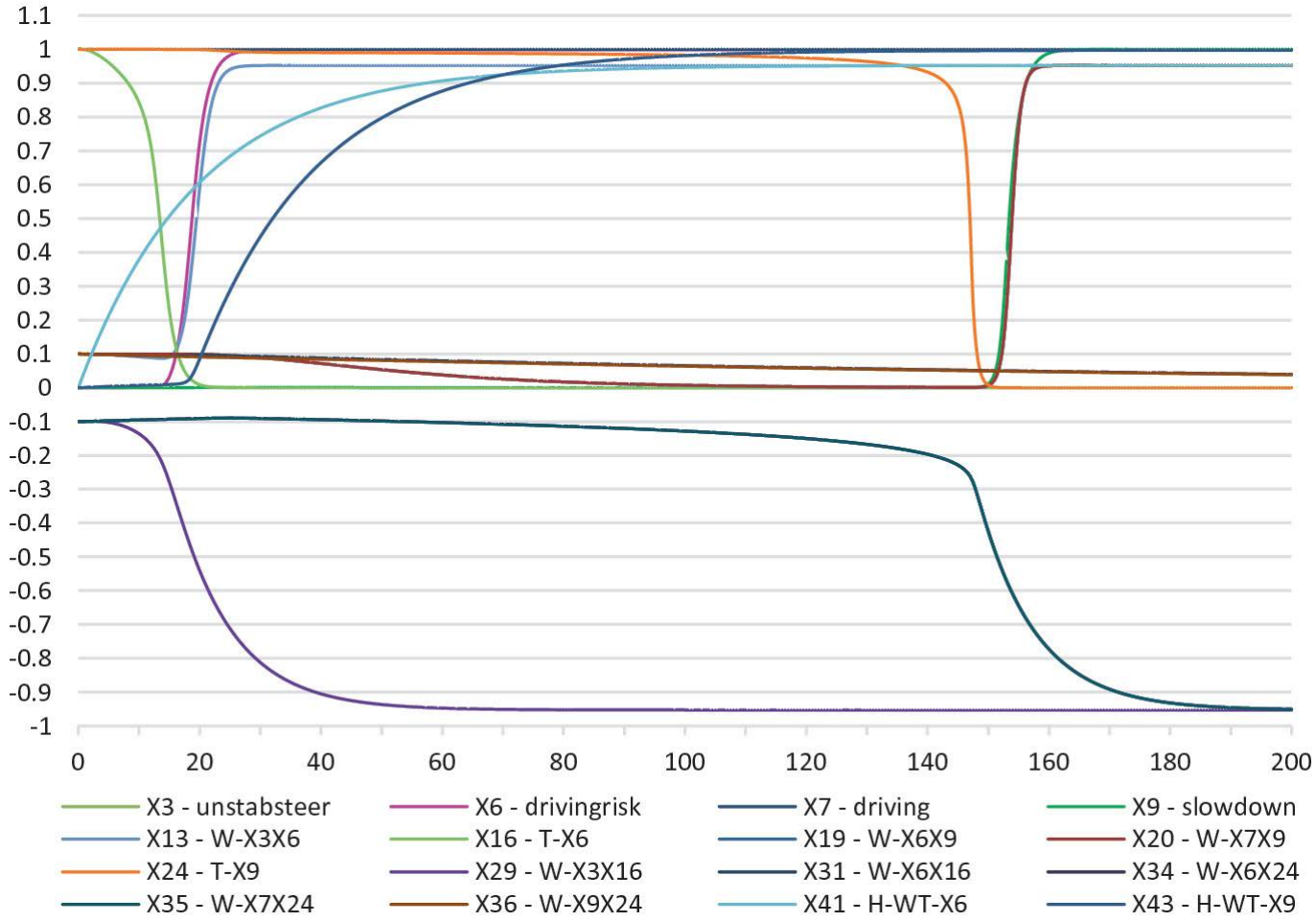


Fig. 7.13 Driving with unstable steering leads to a slow down support action with both adaptive excitability and connectivity

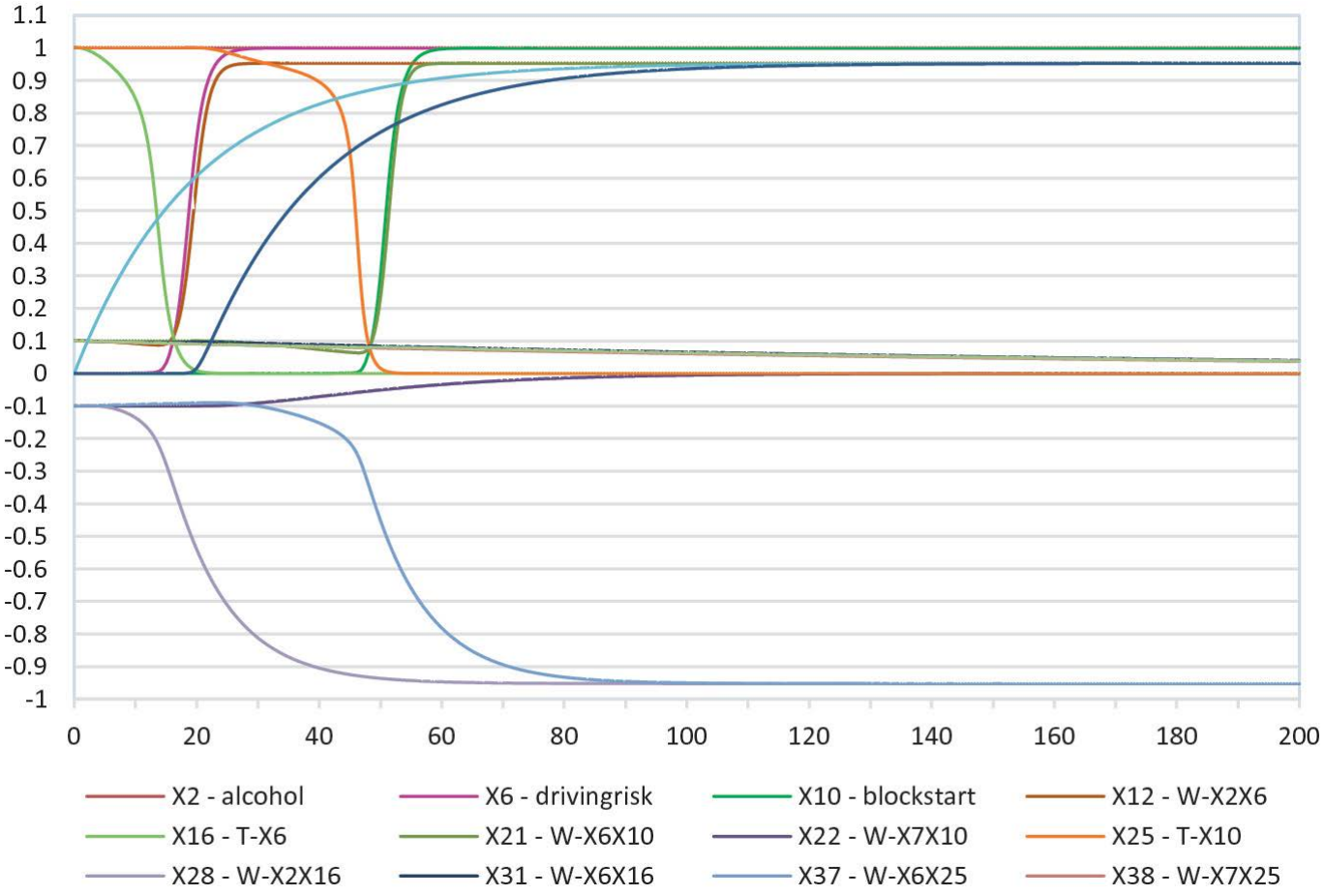


Fig. 7.14 Alcohol usage before intended driving leads to a block start support action with both adaptive excitability and connectivity

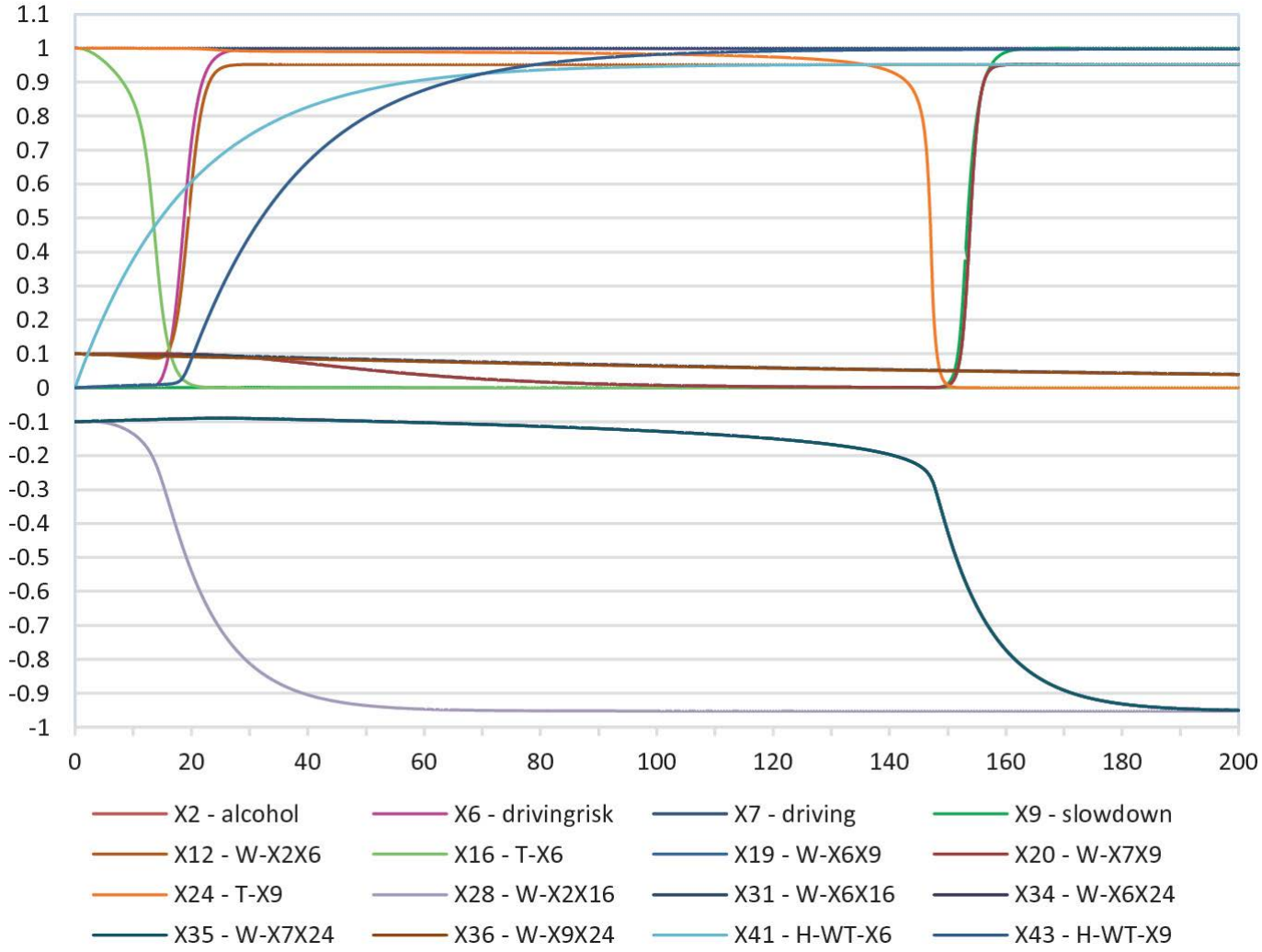


Fig. 7.15 Alcohol usage while driving leads to a slow down support action with both adaptive excitability and connectivity

Scenario 2.1: Driving too long

The first scenario addresses a situation in which the driver drives too long without taking rest. The simulation results of this scenario displayed in Fig. 7.12 show that around time 20 the assessment is made that there is an exhaustion risk (the pink curve) and around time 50 the rest advice support action is generated (the green curve). But before that different forms of adaptation have taken place. First, triggered by the exposure from the context information *longdrive*, the second-order self-model state $\mathbf{H}_{\mathbf{WT}_{\mathrm{exhrisk}}}$ representing the adaptation speed related to the exhaustion risk assessment X_5 starts to increase from 0 (the light blue curve). By this, it exerts its control on adaptation both on the related excitability thresholds and connection weights. The effect of this on adaptation is seen in two ways. Firstly, the adaptation is seen as the decrease of the excitability threshold for the exhaustion risk assessment X_5 (the light green curve going down from 1). Secondly, it is seen as the increase of the connection weight representation X_{11} for the connection from long drive X_1 to exhaustion

risk X_5 (the dark curve going up to around 0.95, hand in hand with the pink curve for the exhaustion risk). These two adaptations together make that the assessment exhaustion risk is generated.

After this, a similar pattern occurs to generate a support action based on the found assessment. This starts around time 20 with the increase of the second-order self-model state $\mathbf{H}_{\mathbf{WT}_{\text{restadvice}}}$ representing the adaptation speed related to the rest advice support action, triggered by the assessment found. By the control exerted by this state, again two types of adaptation take place: the threshold representation X_{23} for rest advice X_8 goes down (orange curve) and both connection weight representations X_{17} and X_{18} go up to around 0.95 (both follow the yellow curve).

Note that behind these first-order learning processes, also a second-order learning process takes place, which is displayed by the two curves for X_{26} and X_{32} that go to around -0.95. Here X_{26} represents the connection weight from X_1 for long drive to X_{15} for the excitability threshold representation for X_5 within the first-order self-model. This exerts a second way of control (in addition to the adaptive first-order adaptation speed) on the first-order adaptation process, as it is this second-order adaptation that enables that the threshold of X_5 gets adapted. Similarly, X_{32} represents the connection weight from X_5 for the exhaustion risk assessment to X_{23} for the excitability threshold representation for X_8 within the first-order self-model. This exerts a second way of control on the first-order adaptation process, as it is this second-order adaptation that enables that the threshold of X_8 gets adapted.

Scenario 2.2: Driving with unstable steering

The second scenario addresses a situation in which the driver shows unstable steering. The simulation results of this scenario displayed in Fig. 7.13 show that around time 20 the assessment is made that there is a driving risk (the pink curve) and around time 160 the slow down support action is generated (the green curve). As in Scenario 2.1, different forms of adaptation have taken place. First, triggered by the exposure from the context information *longdrive*, the second-order self-model state $\mathbf{impact}_{X_k,Y}$ representing the adaptation speed related to the driving risk assessment X_5 starts to increase from 0 (the light blue curve). This exerts its control on adaptation of the related excitability thresholds and connection weights. The first adaptation is seen as the decrease of the excitability threshold for the driving risk assessment X_6 (the light green curve going down from 1). Secondly, it is seen as the increase of the connection weight

representation X_{11} for the connection from unstable steering X_3 to driving risk X_6 (the dark curve going up to around 0.95, hand in hand with the pink curve for the driving risk). These two adaptations together make that the assessment driving risk is generated.

A similar pattern occurs to generate a support action based on the found assessment. This starts around time 25 with the increase of the second-order self-model state $\mathbf{H}_{\mathbf{WT}_{\text{slowdown}}}$ representing the adaptation speed related to the slow down support action, triggered by the assessment found. By the control exerted by this state $\mathbf{H}_{\mathbf{WT}_{\text{slowdown}}}$, again two types of adaptation take place: the threshold representation X_{24} for slow down X_9 goes down (orange curve) and both connection weight representations X_{19} and X_{20} go up to around 0.95 (both follow the yellow curve).

Again, behind these first-order learning processes, also a second-order learning process takes place, which is displayed by the two curves for X_{29} and X_{35} that go to around −0.95. Here X_{29} represents the connection weight from X_3 for unstable steering to X_{16} for the excitability threshold representation for driving risk X_6 within the first-order self-model. This exerts an additional way of control on the first-order adaptation process, as it is this second-order adaptation that enables that the threshold of X_6 gets adapted. Similarly, X_{34} represents the connection weight from X_6 for the driving risk assessment to X_{24} for the excitability threshold representation for X_9 within the first-order self-model. This exerts a second way of control on the first-order adaptation process, as it is this second-order adaptation that enables that the threshold of X_9 gets adapted.

Scenario 2.3: Driving and alcohol

The third scenario addresses a situation in which the driver has taken alcohol. Two cases are considered here (a) the car is not (yet) driving (see Fig. 7.14), and (b) the car is driving (see Fig. 7.15). The simulation results in Fig. 7.14 show that around time 20 the assessment is made that there is a driving risk (the pink curve) and around time 60 the block start support action is generated (the green curve). As in the previous scenarios, different forms of adaptation have taken place; as it goes in a similar manner, this which will not be discussed further for this case.

In Fig. 7.15 the other case is addressed: that the car is driving. Here, a different support action, namely slow down, is generated, which occurs around time 150.

7.7 Discussion

In complex cognitive processes, often internal mental models are used; e.g., (Gentner and Stevens 1983; Greca and Moreira 2000; Kieras and Bovair 1984; Seel 2006). Such models can just be applied, but they are also often adaptive, in order to form and improve them. The focus in this chapter was on adaptive cognitive analysis and support processes for the performance of a human in a demanding task; the adaptive network model was illustrated for a car driver. Within these processes internal mental models are used for the analysis and support processes. Most of the material was adopted from (Treur 2021a, b).

An adaptive network model was presented that models such adaptive cognitive analysis and support processes. The network model makes use of adaptive first-order self-models for the internal mental models used for the analysis and support processes. To control the adaptation of these first-order self-models, second-order self-models are included. In contrast to the model described in (Treur 2021a, b) that only addresses adaptivity of excitability thresholds, in the model presented in the current chapter also adaptivity of all connection weights is addressed. The adaptive network model was illustrated for realistic scenarios for a car driver who gets exhausted, shows unstable steering, shows an unfocused gaze and/or used alcohol.

For the adaptivity and its control, the network model makes use of biologically plausible adaptation principles informed by the Cognitive Neuroscience literature, two within the first-order self-model for adaptation of connectivity and aggregation characteristics of the base network, in particular concerning connection weights and the excitability thresholds (Aizenman and Linden 2000; Chandra and Barkai 2018; Daoudal and Debanne 2003; Debanne et al. 2019; Hebb 1949; Lisman et al. 2018; Titley et al. 2017; Zhang and Linden 2003), and two within the second-order self-model for adaptation of connectivity characteristics (connection weights) and timing characteristics (learning rates) for the first-order self-model by metaplasticity (Abraham and Bear 1996; Garcia 2002; Magerl et al. 2018; Robinson et al. 2016; Sehgal et al. 2013; Sjöström et al. 2008). This study shows how complex adaptive cognitive processes based on internal mental models, including control of adaptation, can be modeled in an adequate manner by multi-order self-modeling networks.

7.8 Appendix: Specification of the Network Model by Role Matrices

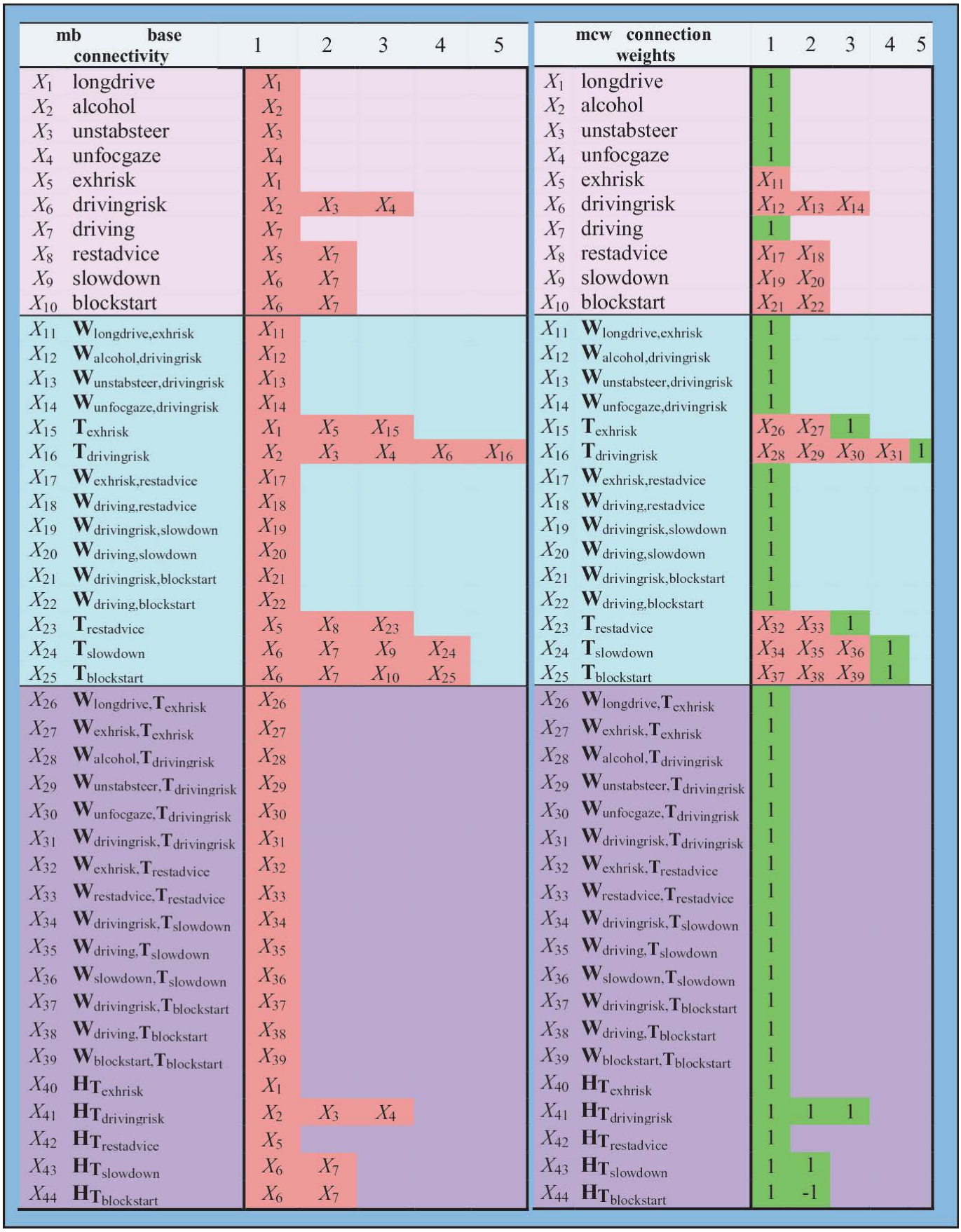
The first network model variant specification concerns a second-order network model where only the excitability thresholds are adaptive, and not the

connection weights. Two combination functions are used, the Euclidean combination function **eucl** and the logistic sum combination function **alogistic** (see Table 7.1).

Role matrices provide an overview of the different types of factors that causally affect the network states. In each of the role matrices, each state X_i in the network has its own row where it is listed which other states or characteristics affect this state from that role. For example, in (base) role matrix **mb** it is indicated which other states affect a given state X_i (because there are incoming connections from them), while in role matrix **mcw** (for connection weights) it is indicated what are the connection weights of these connections. Together, these role matrices **mb** and **mcw** define the *connectivity characteristics* of the network model. Moreover, in role matrix **mcfw** (for combination function weights) it is indicated how a given state X_i is affected by the choice of combination function(s) made for this state, while in role matrix **mcfp** (for combination function parameters) the parameters for these combination functions are indicated. Together, role matrices **mcfw** and **mcfp** define the *aggregation characteristics* of the network model. Finally, the *timing characteristics* of the network model are defined by role matrix **ms** (for speed factors). In the nonempty cells in role matrices there is either a static value or a pointer (reference) to a state that represents this value in a dynamic manner as a self-model state. The latter option is the detailed specification of what in the 3D pictures are the pink downward arrows. It specifies the specific role of the causal effect; this provides a quite compact specification of the different self-model levels.

For the first network model specification, for connectivity characteristics all connection weights not determined by **W**-states are 1, except for the connection from driving to $\mathbf{H}_{\mathbf{WT}_{\text{restadvice}}}$, which is −1. For aggregation characteristics, the logistic sum combination function (see Table 7.1) is used for the base states for assessment and support options (with steepness $\boldsymbol{\sigma} = 8$ and adaptive excitability threshold) and the second-order $\mathbf{H}_{\mathbf{WT}}$-states (with steepness $\boldsymbol{\sigma} = 4$ and excitability threshold $\boldsymbol{\tau} = 0.7$ or 1.4 depending on the number of incoming connections). All other states use the Euclidean combination function (see Table 7.1) with $\boldsymbol{n} = 1$ and $\boldsymbol{\lambda} = 1$, which actually is just a sum function. For timing characteristics, the speed factors of the base states for assessment and support options are 0.5 and for the second-order $\mathbf{H}_{\mathbf{WT}}$-states 0.05. All other speed factors are adaptive (the base states for assessment and support options) or 0 (for the other base states and for all **W**-states). The initial values for all **W**-states (which are kept constant due to the speed factor value 0) are 1 when they represent a positive connection; negative ones are $\mathbf{W}_{\text{driving,blockstart}}$, $\mathbf{W}_{\text{longdrive}}$, $\mathbf{T}_{\text{exhrisk}}$, $\mathbf{W}_{\text{alcohol}}$, $\mathbf{T}_{\text{drivingrisk}}$, $\mathbf{W}_{\text{unstabsteer}}$, $\mathbf{T}_{\text{drivingrisk}}$, $\mathbf{W}_{\text{unfocgaze}}$, $\mathbf{T}_{\text{drivingrisk}}$, $\mathbf{W}_{\text{exhrisk}}$,

$\mathbf{T}_{\text{restadvice}}$ which have initial value −1, and $\mathbf{W}_{\text{drivingrisk}}$, $\mathbf{T}_{\text{slowdown}}$, $\mathbf{W}_{\text{driving}}$, $\mathbf{T}_{\text{slowdown}}$, $\mathbf{W}_{\text{drivingrisk}}$, $\mathbf{T}_{\text{blockstart}}$, $\mathbf{W}_{\text{driving}}$, $\mathbf{T}_{\text{blockstart}}$ with initial value −0.5. The initial values of all H_{WT}-states are 0 as are they for all base states except the observables shown in Table 7.3, which depend on the chosen scenario. Finally, the initial values for the five **T**-states were on purpose set on too high values 2, 1.4, 2.4, 2.8, 2.4, respectively (in relation to the number of their incoming connections), in order to let adaptation happen (Figs. 7.16, 7.17, and 7.18).



mb base connectivity		1	2	3	4	5
X_1	longdrive	X_1				
X_2	alcohol	X_2				
X_3	unstabsteer	X_3				
X_4	unfocgaze	X_4				
X_5	exhrisk	X_1				
X_6	drivingrisk	X_2	X_3	X_4		
X_7	driving	X_7				
X_8	restadvice	X_5	X_7			
X_9	slowdown	X_6	X_7			
X_{10}	blockstart	X_6	X_7			
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	X_{11}				
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	X_{12}				
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	X_{13}				
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	X_{14}				
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	X_1	X_5	X_{15}		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	X_2	X_3	X_4	X_6	X_{16}
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	X_{17}				
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	X_{18}				
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	X_{19}				
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	X_{20}				
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	X_{21}				
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	X_{22}				
X_{23}	$\mathbf{T}_{\text{restadvice}}$	X_5	X_8	X_{23}		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	X_6	X_7	X_9	X_{24}	
X_{25}	$\mathbf{T}_{\text{blockstart}}$	X_6	X_7	X_{10}	X_{25}	
X_{26}	$\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$	X_{26}				
X_{27}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$	X_{27}				
X_{28}	$\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$	X_{28}				
X_{29}	$\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$	X_{29}				
X_{30}	$\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$	X_{30}				
X_{31}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$	X_{31}				
X_{32}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$	X_{32}				
X_{33}	$\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$	X_{33}				
X_{34}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$	X_{34}				
X_{35}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$	X_{35}				
X_{36}	$\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$	X_{36}				
X_{37}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$	X_{37}				
X_{38}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$	X_{38}				
X_{39}	$\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$	X_{39}				
X_{40}	$\mathbf{HT}_{\text{exhrisk}}$	X_1				
X_{41}	$\mathbf{HT}_{\text{drivingrisk}}$	X_2	X_3	X_4		
X_{42}	$\mathbf{HT}_{\text{restadvice}}$	X_5				
X_{43}	$\mathbf{HT}_{\text{slowdown}}$	X_6	X_7			
X_{44}	$\mathbf{HT}_{\text{blockstart}}$	X_6	X_7			

mcw connection weights		1	2	3	4	5
X_1	longdrive	1				
X_2	alcohol	1				
X_3	unstabsteer	1				
X_4	unfocgaze	1				
X_5	exhrisk	X_{11}				
X_6	drivingrisk	X_{12}	X_{13}	X_{14}		
X_7	driving	1				
X_8	restadvice	X_{17}	X_{18}			
X_9	slowdown	X_{19}	X_{20}			
X_{10}	blockstart	X_{21}	X_{22}			
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	1				
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	1				
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	1				
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	1				
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	X_{26}	X_{27}	1		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	X_{28}	X_{29}	X_{30}	X_{31}	1
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	1				
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	1				
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	1				
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	1				
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	1				
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	1				
X_{23}	$\mathbf{T}_{\text{restadvice}}$	X_{32}	X_{33}	1		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	X_{34}	X_{35}	X_{36}	1	
X_{25}	$\mathbf{T}_{\text{blockstart}}$	X_{37}	X_{38}	X_{39}	1	
X_{26}	$\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$	1				
X_{27}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$	1				
X_{28}	$\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$	1				
X_{29}	$\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$	1				
X_{30}	$\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$	1				
X_{31}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$	1				
X_{32}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$	1				
X_{33}	$\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$	1				
X_{34}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$	1				
X_{35}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$	1				
X_{36}	$\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$	1				
X_{37}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$	1				
X_{38}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$	1				
X_{39}	$\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$	1				
X_{40}	$\mathbf{HT}_{\text{exhrisk}}$	1				
X_{41}	$\mathbf{HT}_{\text{drivingrisk}}$	1	1	1		
X_{42}	$\mathbf{HT}_{\text{restadvice}}$	1				
X_{43}	$\mathbf{HT}_{\text{slowdown}}$	1	1			
X_{44}	$\mathbf{HT}_{\text{blockstart}}$	1	-1			

Fig. 7.16 Role matrices specifying *connectivity* characteristics: **mb** for base connections and **mcw** for connection weights

mcfw combination function weights		1 eucl	2 alog-istic
X_1	longdrive	1	
X_2	alcohol	1	
X_3	unstabsteer	1	
X_4	unfocgaze	1	
X_5	exhrisk		1
X_6	drivingrisk		1
X_7	driving	1	
X_8	restadvice		1
X_9	slowdown		1
X_{10}	blockstart		1
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	1	
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	1	
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	1	
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	1	
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	1	
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	1	
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	1	
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	1	
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	1	
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	1	
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	1	
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	1	
X_{23}	$\mathbf{T}_{\text{restadvice}}$	1	
X_{24}	$\mathbf{T}_{\text{slowdown}}$	1	
X_{25}	$\mathbf{T}_{\text{blockstart}}$	1	
X_{26}	$\mathbf{W}_{\text{longdrive},\mathbf{T}_{\text{exhrisk}}}$	1	
X_{27}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{exhrisk}}}$	1	
X_{28}	$\mathbf{W}_{\text{alcohol},\mathbf{T}_{\text{drivingrisk}}}$	1	
X_{29}	$\mathbf{W}_{\text{unstabsteer},\mathbf{T}_{\text{drivingrisk}}}$	1	
X_{30}	$\mathbf{W}_{\text{unfocgaze},\mathbf{T}_{\text{drivingrisk}}}$	1	
X_{31}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{drivingrisk}}}$	1	
X_{32}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{restadvice}}}$	1	
X_{33}	$\mathbf{W}_{\text{restadvice},\mathbf{T}_{\text{restadvice}}}$	1	
X_{34}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{slowdown}}}$	1	
X_{35}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{slowdown}}}$	1	
X_{36}	$\mathbf{W}_{\text{slowdown},\mathbf{T}_{\text{slowdown}}}$	1	
X_{37}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{blockstart}}}$	1	
X_{38}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{blockstart}}}$	1	
X_{39}	$\mathbf{W}_{\text{blockstart},\mathbf{T}_{\text{blockstart}}}$	1	
X_{40}	$\mathbf{HT}_{\text{exhrisk}}$		1
X_{41}	$\mathbf{HT}_{\text{drivingrisk}}$		1
X_{42}	$\mathbf{HT}_{\text{restadvice}}$		1
X_{43}	$\mathbf{HT}_{\text{slowdown}}$		1
X_{44}	$\mathbf{HT}_{\text{blockstart}}$		1

mcfp combination function parameters		1 eucl		2 alogistic	
		1 n	2 λ	1 σ	2 τ
X_1	longdrive	1	1		
X_2	alcohol	1	1		
X_3	unstabsteer	1	1		
X_4	unfocgaze	1	1		
X_5	exhrisk			8	X_{15}
X_6	drivingrisk			8	X_{16}
X_7	driving	1	1		
X_8	restadvice			8	X_{23}
X_9	slowdown			8	X_{24}
X_{10}	blockstart			8	X_{25}
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	1	1		
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	1	1		
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	1	1		
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	1	1		
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	1	1		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	1	1		
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	1	1		
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	1	1		
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	1	1		
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	1	1		
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	1	1		
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	1	1		
X_{23}	$\mathbf{T}_{\text{restadvice}}$	1	1		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	1	1		
X_{25}	$\mathbf{T}_{\text{blockstart}}$	1	1		
X_{26}	$\mathbf{W}_{\text{longdrive},\mathbf{T}_{\text{exhrisk}}}$	1	1		
X_{27}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{exhrisk}}}$	1	1		
X_{28}	$\mathbf{W}_{\text{alcohol},\mathbf{T}_{\text{drivingrisk}}}$	1	1		
X_{29}	$\mathbf{W}_{\text{unstabsteer},\mathbf{T}_{\text{drivingrisk}}}$	1	1		
X_{30}	$\mathbf{W}_{\text{unfocgaze},\mathbf{T}_{\text{drivingrisk}}}$	1	1		
X_{31}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{drivingrisk}}}$	1	1		
X_{32}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{restadvice}}}$	1	1		
X_{33}	$\mathbf{W}_{\text{restadvice},\mathbf{T}_{\text{restadvice}}}$	1	1		
X_{34}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{slowdown}}}$	1	1		
X_{35}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{slowdown}}}$	1	1		
X_{36}	$\mathbf{W}_{\text{slowdown},\mathbf{T}_{\text{slowdown}}}$	1	1		
X_{37}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{blockstart}}}$	1	1		
X_{38}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{blockstart}}}$	1	1		
X_{39}	$\mathbf{W}_{\text{blockstart},\mathbf{T}_{\text{blockstart}}}$	1	1		
X_{40}	$\mathbf{HT}_{\text{exhrisk}}$			4	0.7
X_{41}	$\mathbf{HT}_{\text{drivingrisk}}$			4	0.7
X_{42}	$\mathbf{HT}_{\text{restadvice}}$			4	0.7
X_{43}	$\mathbf{HT}_{\text{slowdown}}$			4	1.4
X_{44}	$\mathbf{HT}_{\text{blockstart}}$			4	0.7

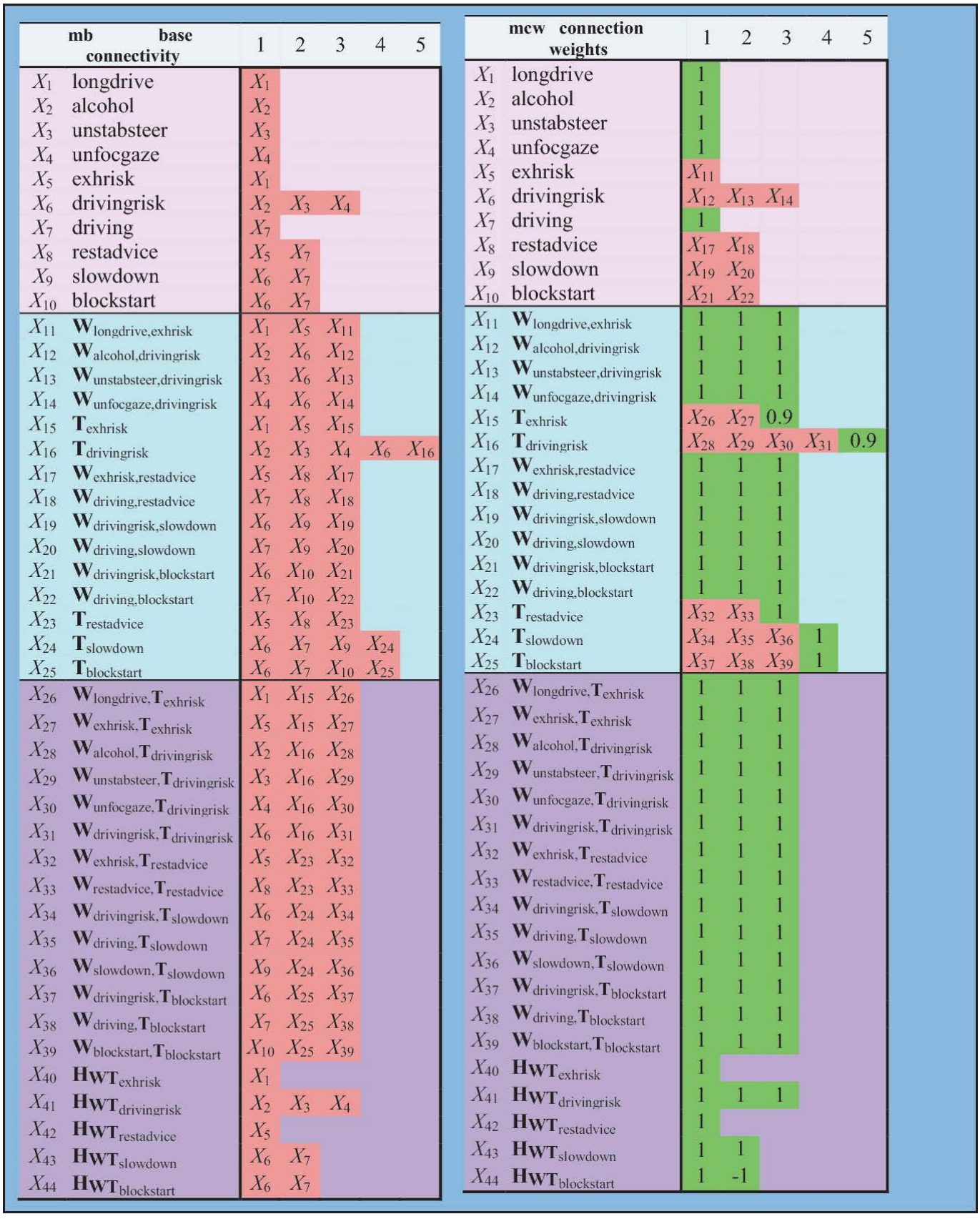
Fig. 7.17 Role matrices specifying *aggregation* characteristics: **mcfw** for combination function weights and **mcfp** for combination function parameters

ms speed factors		1
X_1	longdrive	0
X_2	alcohol	0
X_3	unstabsteer	0
X_4	unfocgaze	0
X_5	exhrisk	0.5
X_6	drivingrisk	0.5
X_7	driving	0
X_8	restadvice	0.5
X_9	slowdown	0.5
X_{10}	blockstart	0.5
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	0
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	0
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	0
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	0
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	X_{40}
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	X_{41}
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	0
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	0
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	0
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	0
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	0
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	0
X_{23}	$\mathbf{T}_{\text{restadvice}}$	X_{42}
X_{24}	$\mathbf{T}_{\text{slowdown}}$	X_{43}
X_{25}	$\mathbf{T}_{\text{blockstart}}$	X_{44}
X_{26}	$\mathbf{W}_{\text{longdrive},\mathbf{T}_{\text{exhrisk}}}$	0
X_{27}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{exhrisk}}}$	0
X_{28}	$\mathbf{W}_{\text{alcohol},\mathbf{T}_{\text{drivingrisk}}}$	0
X_{29}	$\mathbf{W}_{\text{unstabsteer},\mathbf{T}_{\text{drivingrisk}}}$	0
X_{30}	$\mathbf{W}_{\text{unfocgaze},\mathbf{T}_{\text{drivingrisk}}}$	0
X_{31}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{drivingrisk}}}$	0
X_{32}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{restadvice}}}$	0
X_{33}	$\mathbf{W}_{\text{restadvice},\mathbf{T}_{\text{restadvice}}}$	0
X_{34}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{slowdown}}}$	0
X_{35}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{slowdown}}}$	0
X_{36}	$\mathbf{W}_{\text{slowdown},\mathbf{T}_{\text{slowdown}}}$	0
X_{37}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{blockstart}}}$	0
X_{38}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{blockstart}}}$	0
X_{39}	$\mathbf{W}_{\text{blockstart},\mathbf{T}_{\text{blockstart}}}$	0
X_{40}	$\mathbf{HT}_{\text{exhrisk}}$	0.05
X_{41}	$\mathbf{HT}_{\text{drivingrisk}}$	0.05
X_{42}	$\mathbf{HT}_{\text{restadvice}}$	0.05
X_{43}	$\mathbf{HT}_{\text{slowdown}}$	0.05
X_{44}	$\mathbf{HT}_{\text{blockstart}}$	0.05

iv initial values		1
X_1	longdrive	0
X_2	alcohol	0
X_3	unstabsteer	0
X_4	unfocgaze	0
X_5	exhrisk	0
X_6	drivingrisk	0
X_7	driving	0
X_8	restadvice	0
X_9	slowdown	0
X_{10}	blockstart	0
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	1
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	1
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	1
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	1
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	2
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	1.4
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	1
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	1
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	1
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	1
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	1
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	-1
X_{23}	$\mathbf{T}_{\text{restadvice}}$	2.4
X_{24}	$\mathbf{T}_{\text{slowdown}}$	2.8
X_{25}	$\mathbf{T}_{\text{blockstart}}$	2.4
X_{26}	$\mathbf{W}_{\text{longdrive},\mathbf{T}_{\text{exhrisk}}}$	-1
X_{27}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{exhrisk}}}$	1
X_{28}	$\mathbf{W}_{\text{alcohol},\mathbf{T}_{\text{drivingrisk}}}$	-1
X_{29}	$\mathbf{W}_{\text{unstabsteer},\mathbf{T}_{\text{drivingrisk}}}$	-1
X_{30}	$\mathbf{W}_{\text{unfocgaze},\mathbf{T}_{\text{drivingrisk}}}$	-1
X_{31}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{drivingrisk}}}$	1
X_{32}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{restadvice}}}$	-1
X_{33}	$\mathbf{W}_{\text{restadvice},\mathbf{T}_{\text{restadvice}}}$	1
X_{34}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{slowdown}}}$	-0.5
X_{35}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{slowdown}}}$	-0.5
X_{36}	$\mathbf{W}_{\text{slowdown},\mathbf{T}_{\text{slowdown}}}$	1
X_{37}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{blockstart}}}$	-0.5
X_{38}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{blockstart}}}$	-0.5
X_{39}	$\mathbf{W}_{\text{blockstart},\mathbf{T}_{\text{blockstart}}}$	1
X_{40}	$\mathbf{HT}_{\text{exhrisk}}$	0
X_{41}	$\mathbf{HT}_{\text{drivingrisk}}$	0
X_{42}	$\mathbf{HT}_{\text{restadvice}}$	0
X_{43}	$\mathbf{HT}_{\text{slowdown}}$	0
X_{44}	$\mathbf{HT}_{\text{blockstart}}$	0

Fig. 7.18 Role matrix **ms** specifying *timing* characteristics and vector **iv** of initial values

The second network model variant specification concerns a second-order network model where not only the excitability thresholds are adaptive, but also the connection weights (Fig. 7.19).



mb base connectivity		1	2	3	4	5
X_1	longdrive	X_1				
X_2	alcohol	X_2				
X_3	unstabsteer	X_3				
X_4	unfocgaze	X_4				
X_5	exhrisk	X_1				
X_6	drivingrisk	X_2	X_3	X_4		
X_7	driving	X_7				
X_8	restadvice	X_5	X_7			
X_9	slowdown	X_6	X_7			
X_{10}	blockstart	X_6	X_7			
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	X_1	X_5	X_{11}		
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	X_2	X_6	X_{12}		
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	X_3	X_6	X_{13}		
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	X_4	X_6	X_{14}		
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	X_1	X_5	X_{15}		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	X_2	X_3	X_4	X_6	X_{16}
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	X_5	X_8	X_{17}		
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	X_7	X_8	X_{18}		
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	X_6	X_9	X_{19}		
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	X_7	X_9	X_{20}		
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	X_6	X_{10}	X_{21}		
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	X_7	X_{10}	X_{22}		
X_{23}	$\mathbf{T}_{\text{restadvice}}$	X_5	X_8	X_{23}		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	X_6	X_7	X_9	X_{24}	
X_{25}	$\mathbf{T}_{\text{blockstart}}$	X_6	X_7	X_{10}	X_{25}	
X_{26}	$\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$	X_1	X_{15}	X_{26}		
X_{27}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$	X_5	X_{15}	X_{27}		
X_{28}	$\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$	X_2	X_{16}	X_{28}		
X_{29}	$\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$	X_3	X_{16}	X_{29}		
X_{30}	$\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$	X_4	X_{16}	X_{30}		
X_{31}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$	X_6	X_{16}	X_{31}		
X_{32}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$	X_5	X_{23}	X_{32}		
X_{33}	$\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$	X_8	X_{23}	X_{33}		
X_{34}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$	X_6	X_{24}	X_{34}		
X_{35}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$	X_7	X_{24}	X_{35}		
X_{36}	$\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$	X_9	X_{24}	X_{36}		
X_{37}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$	X_6	X_{25}	X_{37}		
X_{38}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$	X_7	X_{25}	X_{38}		
X_{39}	$\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$	X_{10}	X_{25}	X_{39}		
X_{40}	$\mathbf{HWT}_{\text{exhrisk}}$	X_1				
X_{41}	$\mathbf{HWT}_{\text{drivingrisk}}$	X_2	X_3	X_4		
X_{42}	$\mathbf{HWT}_{\text{restadvice}}$	X_5				
X_{43}	$\mathbf{HWT}_{\text{slowdown}}$	X_6	X_7			
X_{44}	$\mathbf{HWT}_{\text{blockstart}}$	X_6	X_7			

mcw connection weights		1	2	3	4	5
X_1	longdrive	1				
X_2	alcohol	1				
X_3	unstabsteer	1				
X_4	unfocgaze	1				
X_5	exhrisk	X_{11}				
X_6	drivingrisk	X_{12}	X_{13}	X_{14}		
X_7	driving	1				
X_8	restadvice	X_{17}	X_{18}			
X_9	slowdown	X_{19}	X_{20}			
X_{10}	blockstart	X_{21}	X_{22}			
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	1	1	1		
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	1	1	1		
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	1	1	1		
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	1	1	1		
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	X_{26}	X_{27}	0.9		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	X_{28}	X_{29}	X_{30}	X_{31}	0.9
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	1	1	1		
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	1	1	1		
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	1	1	1		
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	1	1	1		
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	1	1	1		
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	1	1	1		
X_{23}	$\mathbf{T}_{\text{restadvice}}$	X_{32}	X_{33}	1		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	X_{34}	X_{35}	X_{36}	1	
X_{25}	$\mathbf{T}_{\text{blockstart}}$	X_{37}	X_{38}	X_{39}	1	
X_{26}	$\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$	1	1	1		
X_{27}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$	1	1	1		
X_{28}	$\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$	1	1	1		
X_{29}	$\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$	1	1	1		
X_{30}	$\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$	1	1	1		
X_{31}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$	1	1	1		
X_{32}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$	1	1	1		
X_{33}	$\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$	1	1	1		
X_{34}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$	1	1	1		
X_{35}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$	1	1	1		
X_{36}	$\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$	1	1	1		
X_{37}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$	1	1	1		
X_{38}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$	1	1	1		
X_{39}	$\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$	1	1	1		
X_{40}	$\mathbf{HWT}_{\text{exhrisk}}$	1				
X_{41}	$\mathbf{HWT}_{\text{drivingrisk}}$	1	1	1		
X_{42}	$\mathbf{HWT}_{\text{restadvice}}$	1				
X_{43}	$\mathbf{HWT}_{\text{slowdown}}$	1	1			
X_{44}	$\mathbf{HWT}_{\text{blockstart}}$	1	-1			

Fig. 7.19 Role matrices specifying *connectivity* characteristics: **mb** for base connections and **mcw** for connection weights

Three combination functions are used, the logistic sum combination function **alogistic**, and the hebbian learning functions **hebb** and **hebbneg** (see Tables 7.1 and 7.2). For a general explanation of role matrices, see Appendix A (Figs. 7.20 and 7.21).

mcfw combination function weights		1 alog-istic	2 hebb	3 hebb neg
X_1	longdrive	1		
X_2	alcohol	1		
X_3	unstabsteer	1		
X_4	unfocgaze	1		
X_5	exhrisk	1		
X_6	drivingrisk	1		
X_7	driving	1		
X_8	restadvice	1		
X_9	slowdown	1		
X_{10}	blockstart	1		
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$		1	
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$		1	
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$		1	
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$		1	
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	1		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	1		
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$		1	
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$		1	
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$		1	
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$		1	
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$		1	
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$			1
X_{23}	$\mathbf{T}_{\text{restadvice}}$	1		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	1		
X_{25}	$\mathbf{T}_{\text{blockstart}}$	1		
X_{26}	$\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$			1
X_{27}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$		1	
X_{28}	$\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$			1
X_{29}	$\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$			1
X_{30}	$\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$			1
X_{31}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$		1	
X_{32}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$			1
X_{33}	$\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$		1	
X_{34}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$			1
X_{35}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$			1
X_{36}	$\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$		1	
X_{37}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$			1
X_{38}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$		1	
X_{39}	$\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$		1	
X_{40}	$\mathbf{HWT}_{\text{exhrisk}}$	1		
X_{41}	$\mathbf{HWT}_{\text{drivingrisk}}$	1		
X_{42}	$\mathbf{HWT}_{\text{restadvice}}$	1		
X_{43}	$\mathbf{HWT}_{\text{slowdown}}$	1		
X_{44}	$\mathbf{HWT}_{\text{blockstart}}$	1		

mcfp combination function parameters		1 alogistic		2 hebb	3 hebbneg
		1	2	1	1
		n	τ	μ	μ
X_1	longdrive	8	0.5		
X_2	alcohol	8	0.5		
X_3	unstabsteer	8	0.5		
X_4	unfocgaze	8	0.5		
X_5	exhrisk	8	X_{15}		
X_6	drivingrisk	8	X_{16}		
X_7	driving	8	0.5		
X_8	restadvice	8	X_{23}		
X_9	slowdown	8	X_{24}		
X_{10}	blockstart	8	X_{25}		
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$			0.95	
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$			0.95	
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$			0.95	
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$			0.95	
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	8	0.5		
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	8	0.5		
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$			0.95	
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$			0.95	
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$			0.95	
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$			0.95	
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$			0.95	
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$				0.95
X_{23}	$\mathbf{T}_{\text{restadvice}}$	8	0.5		
X_{24}	$\mathbf{T}_{\text{slowdown}}$	8	0.2		
X_{25}	$\mathbf{T}_{\text{blockstart}}$	8	0.5		
X_{26}	$\mathbf{W}_{\text{longdrive,}}\mathbf{T}_{\text{exhrisk}}$				0.95
X_{27}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{exhrisk}}$			0.95	
X_{28}	$\mathbf{W}_{\text{alcohol,}}\mathbf{T}_{\text{drivingrisk}}$				0.95
X_{29}	$\mathbf{W}_{\text{unstabsteer,}}\mathbf{T}_{\text{drivingrisk}}$				0.95
X_{30}	$\mathbf{W}_{\text{unfocgaze,}}\mathbf{T}_{\text{drivingrisk}}$				0.95
X_{31}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{drivingrisk}}$			0.95	
X_{32}	$\mathbf{W}_{\text{exhrisk,}}\mathbf{T}_{\text{restadvice}}$				0.95
X_{33}	$\mathbf{W}_{\text{restadvice,}}\mathbf{T}_{\text{restadvice}}$			0.95	
X_{34}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{slowdown}}$				0.95
X_{35}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{slowdown}}$				0.95
X_{36}	$\mathbf{W}_{\text{slowdown,}}\mathbf{T}_{\text{slowdown}}$			0.95	
X_{37}	$\mathbf{W}_{\text{drivingrisk,}}\mathbf{T}_{\text{blockstart}}$				0.95
X_{38}	$\mathbf{W}_{\text{driving,}}\mathbf{T}_{\text{blockstart}}$			0.95	
X_{39}	$\mathbf{W}_{\text{blockstart,}}\mathbf{T}_{\text{blockstart}}$			0.95	
X_{40}	$\mathbf{HWT}_{\text{exhrisk}}$	10	0.7		
X_{41}	$\mathbf{HWT}_{\text{drivingrisk}}$	10	0.7		
X_{42}	$\mathbf{HWT}_{\text{restadvice}}$	10	0.7		
X_{43}	$\mathbf{HWT}_{\text{slowdown}}$	10	1.4		
X_{44}	$\mathbf{HWT}_{\text{blockstart}}$	10	0.7		

Fig. 7.20 Role matrices specifying *aggregation* characteristics: **mcfw** for combination function weights and **mcfp** for combination function parameters

ms speed factors		1
X_1	longdrive	0
X_2	alcohol	0
X_3	unstabsteer	0
X_4	unfocgaze	0
X_5	exhrisk	0.5
X_6	drivingrisk	0.5
X_7	driving	0
X_8	restadvice	0.5
X_9	slowdown	0.5
X_{10}	blockstart	0.5
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	X_{40}
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	X_{41}
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	X_{41}
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	X_{41}
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	X_{40}
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	X_{41}
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	X_{42}
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	X_{42}
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	X_{43}
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	X_{43}
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	X_{44}
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	X_{44}
X_{23}	$\mathbf{T}_{\text{restadvice}}$	X_{42}
X_{24}	$\mathbf{T}_{\text{slowdown}}$	X_{43}
X_{25}	$\mathbf{T}_{\text{blockstart}}$	X_{44}
X_{26}	$\mathbf{W}_{\text{longdrive},\mathbf{T}_{\text{exhrisk}}}$	0.1
X_{27}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{exhrisk}}}$	0.1
X_{28}	$\mathbf{W}_{\text{alcohol},\mathbf{T}_{\text{drivingrisk}}}$	0.1
X_{29}	$\mathbf{W}_{\text{unstabsteer},\mathbf{T}_{\text{drivingrisk}}}$	0.1
X_{30}	$\mathbf{W}_{\text{unfocgaze},\mathbf{T}_{\text{drivingrisk}}}$	0.1
X_{31}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{drivingrisk}}}$	0.1
X_{32}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{restadvice}}}$	0.1
X_{33}	$\mathbf{W}_{\text{restadvice},\mathbf{T}_{\text{restadvice}}}$	0.1
X_{34}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{slowdown}}}$	0.1
X_{35}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{slowdown}}}$	0.1
X_{36}	$\mathbf{W}_{\text{slowdown},\mathbf{T}_{\text{slowdown}}}$	0.1
X_{37}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{blockstart}}}$	0.1
X_{38}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{blockstart}}}$	0.1
X_{39}	$\mathbf{W}_{\text{blockstart},\mathbf{T}_{\text{blockstart}}}$	0.1
X_{40}	$\mathbf{HWT}_{\text{exhrisk}}$	0.05
X_{41}	$\mathbf{HWT}_{\text{drivingrisk}}$	0.05
X_{42}	$\mathbf{HWT}_{\text{restadvice}}$	0.05
X_{43}	$\mathbf{HWT}_{\text{slowdown}}$	0.05
X_{44}	$\mathbf{HWT}_{\text{blockstart}}$	0.05

iv initial values		1
X_1	longdrive	0
X_2	alcohol	0
X_3	unstabsteer	0
X_4	unfocgaze	0
X_5	exhrisk	0
X_6	drivingrisk	0
X_7	driving	0
X_8	restadvice	0
X_9	slowdown	0
X_{10}	blockstart	0
X_{11}	$\mathbf{W}_{\text{longdrive,exhrisk}}$	0.1
X_{12}	$\mathbf{W}_{\text{alcohol,drivingrisk}}$	0.1
X_{13}	$\mathbf{W}_{\text{unstabsteer,drivingrisk}}$	0.1
X_{14}	$\mathbf{W}_{\text{unfocgaze,drivingrisk}}$	0.1
X_{15}	$\mathbf{T}_{\text{exhrisk}}$	1
X_{16}	$\mathbf{T}_{\text{drivingrisk}}$	1
X_{17}	$\mathbf{W}_{\text{exhrisk,restadvice}}$	0.1
X_{18}	$\mathbf{W}_{\text{driving,restadvice}}$	0.1
X_{19}	$\mathbf{W}_{\text{drivingrisk,slowdown}}$	0.1
X_{20}	$\mathbf{W}_{\text{driving,slowdown}}$	0.1
X_{21}	$\mathbf{W}_{\text{drivingrisk,blockstart}}$	0.1
X_{22}	$\mathbf{W}_{\text{driving,blockstart}}$	-0.1
X_{23}	$\mathbf{T}_{\text{restadvice}}$	1
X_{24}	$\mathbf{T}_{\text{slowdown}}$	1
X_{25}	$\mathbf{T}_{\text{blockstart}}$	1
X_{26}	$\mathbf{W}_{\text{longdrive},\mathbf{T}_{\text{exhrisk}}}$	-0.1
X_{27}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{exhrisk}}}$	0.1
X_{28}	$\mathbf{W}_{\text{alcohol},\mathbf{T}_{\text{drivingrisk}}}$	-0.1
X_{29}	$\mathbf{W}_{\text{unstabsteer},\mathbf{T}_{\text{drivingrisk}}}$	-0.1
X_{30}	$\mathbf{W}_{\text{unfocgaze},\mathbf{T}_{\text{drivingrisk}}}$	-0.1
X_{31}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{drivingrisk}}}$	0.1
X_{32}	$\mathbf{W}_{\text{exhrisk},\mathbf{T}_{\text{restadvice}}}$	-0.1
X_{33}	$\mathbf{W}_{\text{restadvice},\mathbf{T}_{\text{restadvice}}}$	0.1
X_{34}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{slowdown}}}$	-0.1
X_{35}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{slowdown}}}$	-0.1
X_{36}	$\mathbf{W}_{\text{slowdown},\mathbf{T}_{\text{slowdown}}}$	0.1
X_{37}	$\mathbf{W}_{\text{drivingrisk},\mathbf{T}_{\text{blockstart}}}$	-0.1
X_{38}	$\mathbf{W}_{\text{driving},\mathbf{T}_{\text{blockstart}}}$	0.1
X_{39}	$\mathbf{W}_{\text{blockstart},\mathbf{T}_{\text{blockstart}}}$	0.1
X_{40}	$\mathbf{HWT}_{\text{exhrisk}}$	0
X_{41}	$\mathbf{HWT}_{\text{drivingrisk}}$	0
X_{42}	$\mathbf{HWT}_{\text{restadvice}}$	0
X_{43}	$\mathbf{HWT}_{\text{slowdown}}$	0
X_{44}	$\mathbf{HWT}_{\text{blockstart}}$	0

Fig. 7.21 Role matrix **ms** specifying *timing* characteristics and vector **iv** of initial values

References

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

Aizenman, C.D., Linden, D.J.: Rapid, synaptically driven increases in the intrinsic excitability of cerebellar deep nuclear neurons. Nat. Neurosci. **3**, 109–111 (2000)
[Crossref]

Bhalwankar, R., Treur, J.: Modeling the Development of Internal Mental Models by an Adaptive Network Model. Proceedings of the 11th Annual International Conference on Brain-Inspired Cognitive Architectures for AI, BICA*AI'20. Procedia Computer Science, Elsevier (2021)

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

Daoudal, G., Debanne, D.: Long-term plasticity of intrinsic excitability: learning rules and mechanisms. Learn. Mem. **10**, 456–465 (2003)
[Crossref]

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

Garcia, R.: Stress, metaplasticity, and antidepressants. Curr. Mol. Med. **2**, 629–638 (2002)
[Crossref]

Gentner, D., Stevens, A.L.: Mental Models. Erlbaum, Hillsdale NJ (1983)

Greca, I.M., Moreira, M.A.: Mental models, conceptual models, and modelling. Int. J. Sci. Educ. **22**(1), 1–11 (2000)
[Crossref]

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

Keysers, C., Gazzola, V.: Hebbian learning and predictive mirror neurons for actions, sensations and emotions. Philos. Trans. r. Soc. Lond. B Biol. Sci. **369**, 20130175 (2014)
[Crossref]

Kieras, D.E., Bovair, S.: The role of a mental model in learning to operate a device. Cogn. Sci. **8**(3), 255–273 (1984)
[Crossref]

Lisman, J., Cooper, K., Sehgal, M., Silva, A.J.: Memory formation depends on both synapse-specific modifications of synaptic strength and cell-specific increases in excitability. Nat. Neurosci. **21**, 309–314 (2018)
[Crossref]

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)
[Crossref]

Robinson, B.L., Harper, N.S., McAlpine, D.: Meta-adaptation in the auditory midbrain under cortical influence. Nat. Commun. **7**, 13442 (2016)
[Crossref]

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

Sehgal, M., Song, C., Ehlers, V.L., Moyer, J.R., Jr.: Learning to learn—Intrinsic plasticity as a metaplasticity mechanism for memory formation. Neurobiol. Learn. Mem. **105**, 186–199 (2013)

Shatz, C.J.: The developing brain. Sci. Am. **267**, 60–67 (1992). https://doi.org/10.1038/scientificamerican0992-60

Sjöström, P.J., Rancz, E.A., Roth, A., Hausser, M.: Dendritic Excitability and Synaptic Plasticity. Physiol Rev **88**, 769–840 (2008)
[Crossref]

Titley, H.K., Brunel, N., Hansel, C.: Toward a neurocentric view of learning. Neuron **95**, 19–32 (2017)
[Crossref]

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive. Springer Publishers, Affective and Social Interactions (2016)

Treur, J.: Modeling higher-order adaptivity of a network by multilevel network reification. Netw. Sci. **8**, S110–S144 (2020a)
[Crossref]

Treur, J.: Network-Oriented Modeling for Adaptive Networks: Designing Higher-order Adaptive Biological, Mental and Social Network Models. Springer Nature Publishing, Cham, Switzerland (2020b)
[Crossref]

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 & Their Applications IX. Proceedings COMPLEX NETWORKS 2020. Studies in Computational Intelligence, vol. 944, pp. 260–274. Springer Nature Switzerland AG (2021a)

Treur, J.: A self-modeling network model addressing controlled adaptive mental models for analysis and support processes. Complex Syst. J. **30**(4), 483–512 (2021b)

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

Zhang, W., Linden, D.J.: The other side of the engram: experience-driven changes in neuronal intrinsic excitability. Nat. Rev. Neurosci. **4**, 885–900 (2003)
[Crossref]

J. Treur, L. Van Ments (eds.), *Mental Models and Their Dynamics, Adaptation, and Control*, Studies in Systems, Decision and Control 394
https://doi.org/10.1007/978-3-030-85821-6_8

8. Who Am I Really: An Adaptive Network Model Addressing Mental Models for Self-referencing, Self-awareness and Self-interpretation

Jan Treur[1] ✉ and Gerrit Glas[2, 3] ✉
(1) Social AI Group, Department of Computer Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
(2) Epistemology and Metaphysics Group, Department of Philosophy, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
(3) Anatomy and Neurosciences Department, Amsterdam University Medical Center, Amsterdam, The Netherlands

✉ **Jan Treur (Corresponding author)**
Email: j.treur@vu.nl

✉ **Gerrit Glas**
Email: g.glas@vu.nl

Abstract
In this chapter, a multilevel cognitive architecture is introduced that can be used to model mental processes in clients of psychotherapeutic sessions and in particular the mental (self-)models they have about themselves. The architecture does not only cover base level mental processes but also mental processes involving self-referencing, self-awareness and self-interpretation. To this end, the cognitive architecture was designed according to four levels, where (part of) the structure of each level is represented by an explicit self-model of it at the next-higher level of the architecture. At that next-higher level, states represent part of the structure of the level below; these states have a referencing relation to it. In this way the overall architecture includes its own overall self-model. The cognitive architecture was evaluated for a case study of a realistic type of therapeutic session from clinical practice.

Keywords Network model – Third-order adaptive – Self-referencing – Self-awareness – Self-interpretation

8.1 Introduction

Within the discipline of Psychiatry there is a longstanding tradition and much experience in supporting clients in the path toward discovery of themselves. Many types of therapy at least partly aim at enhancing the client's knowledge and awareness of her or himself and use this to his or her benefit. This may involve elements such as:

- Getting familiar with one's own personal characteristics
- Being aware of important aspects of oneself
- Being able to interpret one's behaviour in relation to other aspects of oneself
- Based on such insights being able to manage oneself more effectively.

Much of the knowledge about such therapeutic processes has been acquired within a practical clinical context. This can be considered a very valuable source of knowledge which only very partially has been exploited and further developed from a more academic research perspective; see however (Montgomery 2006; Nicolini 2012; Gascoigne and Thornton 2014). So far, to our knowledge, no form of formalisation has been applied or any further computational analysis been executed, except for statistical analyses of gathered data.

The current chapter does address formalisation and more detailed computational analysis of the types of processes and mental self-models as briefly sketched above. Based on a recent conceptual analysis of clinical practice (Glas 2017, 2019), a cognitive architecture has been designed involving the relevant processes of self-referencing, self-awareness and self-interpretation. This cognitive architecture has been specified formally using the network-oriented modeling language described in more detail in (Treur 2020b). Based on this specification, for a particular case study simulation experiments have been conducted using the available dedicated software environment.

In the chapter, first in Sect. 8.2 the psychiatric context addressed is briefly outlined and describes the case study addressed in later sections. In Sect. 8.3 the notion of self-modeling network is briefly discussed. Section 8.4 shows how these notions are useful to model the relevant elements of the considered psychiatric context and discusses a global overview of the designed multilevel cognitive architecture. In Sect. 8.5 an example case study is used to illustrate the use of the architecture, while Sect. 8.6 provides a detailed specification of this. Section 8.7 presents an example simulation for this case study. Finally, Sect. 8.8 is a discussion.

8.2 Perspectives from a Psychiatric Context

This section briefly explores the notions self-referentiality, self-awareness and self-interpretation that play an important role in clinical practice and are the basis of the cognitive architecture introduced in next sections.

8.2.1 Self-referentiality

Many symptoms in the context of psychiatric illness are not just expressions of an underlying molecular, physiological or even psychological derailment but also implicitly refer to the person having them. Strikingly enough, academic psychology and psychiatry have largely ignored this obvious fact. They still lack a conceptual framework to make sense of the self-referencing quality of the emotions, gestures, verbal expressions, and interactions that occur in the context of mental illness. The term self-referentiality is meant to highlight the implicit signifying aspect of these behaviors. These behaviours refer to a 'self', i.e., to an aspect or to aspects of the person having them.

In what follows we will restrict the term to the *implicit* signifying aspects of affective, communicative, verbal and social behaviors. Self-referentiality is not adopting a stance toward these behaviors. Adopting a stance is a form of self-relating. Self-referentiality differs from self-relating in that the behaviors just mentioned by themselves 'say' (signify, indicate) something about the person having them. Their referring quality is not a product of self-reflection and interpretation; the phenomena themselves signify and reveal what is going on with the person having them. This self-referential signifying does not exclude self-awareness, self-relating or self-interpretation. Emotions, for instance, don't lose their self-referential qualities when one becomes aware of them. Self-referentiality and self-awareness may go together. But it still makes sense to distinguish them conceptually; see also (Glas 2017, p. 146).

8.2.2 Self-awareness

By self-awareness we mean that there exists an awareness of an aspect of oneself or of oneself as a whole. It is a state of mind in which an aspect of oneself, or one's self, becomes the object of consciousness (or: reflection); e.g., (Glas 2017, p. 146).

8.2.3 Self-interpretation

Self-interpretation refers to the way people understand (perceive, value) themselves. Self-interpretation builds forth on self-awareness and on self-referential aspects of certain behaviors. It often leads to a re-evaluation of one's initial perceptions. Self-interpretation is needed when the initial awareness of an expression, thought, or utterance is unclear or ambiguous. I may feel hurt by someone without initially knowing why. Upon further reflection, I may discover

why I felt hurt. I may realize how the other person subtly awakened my latent feelings of inferiority or threatened my feelings about someone I love. The anger again reveals that something that (or: someone who) is important for me, is threatened. Further reflection (or discussion with others) makes me aware of the nature and severity of the threat; e.g., (Glas, 2019, p. 29).

8.2.4 Other Literature

We understand that the conceptual framework we are describing looks novel and that it does not seem to neatly fit within the current scene of cognitive science, psychopathology and even philosophy. To be honest, the term 'self' has become important in current psychology and cognitive neuroscience. There exist, for instance, a lot of research on emotion-regulation and self-regulation (Gross and Thompson 2007; Gross and Jazaieri 2014; Tracy et al. 2014). While the concepts of emotion-regulation and self-regulation are evidently highly relevant for psychiatry, they typically aim at what occurs *with* emotions, i.e., post hoc. Koole (2009) defines emotion-regulation as "the set of processes whereby people seek to redirect the spontaneous flow of their emotions". Gross (2008) states that emotion regulation is about "how we try to influence which emotions we have, when we have them, and how we experience and express these emotions"; see also (Gross and Thompson 2007) and for an opposite view, (Kappas 2011).

The concept of the self is an emerging theme in many other respects, i.e., in the context of cognitive neuroscience (Christoff et al. 2011; Damasio 1999, 2010; Immordino-Yang 2011; Northoff et al. 2011; Reddy 2009), developmental psychology (Fonagy et al. 2002; Hobson 2010), general psychology and personality theory (Leary and Tangney 2003), social psychology (Tracy and Robins 2007), and philosophy (Gallagher 2013; Metzinger 2003). However, the focus of attention has so far often been on so-called self-conscious and moral emotions like shame, guilt, embarrassment, social anxiety, pride, and ambivalence (Leary 2007; Prinz 2010; Rorty 2010). The self-conscious emotions have the self as focus or object, though usually indirectly, by making inferences about other people's evaluations of oneself. With self-referentiality we mean something else than this kind of self-concern, however.

Similar accounts are rare, indeed, and are to be found in the phenomenological tradition (Solomon 1983; Stern 2004; Zahavi 2005; Ratcliffe 2005, 2008, 2010; Slaby and Stephan 2008; Stephan 2012; Atkinson and Ratcliffe 2012). The term self-referentiality (and: self-referencing) was initially coined by Paul Ricoeur (1992). The term indicated the implicit reflexivity (or the 'ipse') of symbolic expressions. Our account on self-referentiality comes closest to Ricoeur's; however, he seems to include forms of referring that are part of one's conscious awareness and, therefore, are ready to be interpreted. So, we stick to a

more restrictive definition of self-referencing, i.e., as an implicit form of signification.

8.2.5 Point of Departure for the Case Study Used

In this article we use a simple example to illustrate the architecture that is indicated with the terms self-referentiality, self-awareness, and self-interpretation. We depart with a slightly changed and extended variant of a case description that was given in another work by one of the authors (Glas 2017, p. 151):

> John is a 24 year old student, with a positive family history for mood disorder. He has recently developed a moderately severe depression. John has some physical complaints related to the depression and visits his general practitioner who assures him that there is nothing wrong physically. The GP considers psychosocial stress factors as a possible source of what he calls 'the burn-out'. John tells that he broke up with a girlfriend six months ago and that he has difficulty with finishing his master thesis. However, he has broken up earlier with other girlfriends and he does not seem very worried about the lack of progress with his thesis. The GP suggests physical exercises and prescribes sleeping pills for a period of two weeks, after which he wants to see John again. John does not show-up, however, and his condition worsens. Finishing his study doesn't make sense. He doesn't have a future and the world would be better off without him, he thinks. His friends try to motivate him to look for professional help but John refuses. He doesn't do his physical exercises, begins to drink large amounts of alcohol and slips into a state of sheer passivity.

8.3 Self-modeling Network Models

In the previous section it may already have become clear conceptually and linguistically that for their content many of the concepts used have some reference relation to other concepts. For example, self-awareness refers to the content of the awareness which can be one of oneself's personal characteristics, and in turn such a characteristic refers to its content which can be a relation between mental states such as a memory and a belief. To formalise such reference relations, the notion of self-modeling network is useful. Within AI and related disciplines sometimes the term reification is used for this; e.g., (Galton 2006). Similar, related concepts within AI and Computer Science are reflective architectures, metalevel architectures, meta-interpreters, and metaprogramming; e.g., (Bowen and Kowalski 1982; Sterling and Beer 1989;

Weyhrauch 1980). In this section the notion of self-modeling network is discussed in some more detail and it is indicated how it has been used in (Treur 2020b) for network-oriented modeling in an iterative manner, thus obtaining multilevel self-modeling network models. Such network models will turn out to provide an adequate basis to formalise what was described in Sect. 8.2.

8.3.1 Using Self-models Within a Network Model

We often describe the dynamics of processes in any domain of the world by causal relations, and certainly this usually happens when a biological perspective is used. In principle, the occurrences of these causal relations are based on the configuration of the world; this configuration by itself can be dynamic as well, which is usually described as adaptation. For example:

(a) Dynamics of neural or mental states based on the causal relations as used to describe the brain; e.g., (Kim 1996). Adaptation of these causal relations occurs, for example, as changing synapses or excitation thresholds within the brain; e.g., (Hebb 1949; Tse 2013; Chandra and Barkai 2018)

(b) Dynamics of social processes based on causal relations describing how individuals affect each other as used within the social domain; e.g., (Levy and Nail 1993; Iacoboni 2008). Adaptation of these causal relations within the social domain, occurs as changing connections by bonding and within the underlying brain processes such as mirroring; e.g., (McPherson et al. 2001; Iacoboni 2008; Keysers and Gazzola 2014)

(c) Dynamics of biochemical processes based on the causal relations as used within the biochemical domain. Adaptation of these causal relations occurs as changes in such networks; e.g., (Westerhoff et al. 2014a, b)

(d) Dynamics due to causal pathways within organisms (Westerhoff et al. 2014a, b). Adaptation within evolutionary biological processes occurs as changing the causal pathways; e.g., (Fessler et al. 2005, 2015)

(e) Dynamics of physical processes based on causal relations as used within the physical domain. Adaptation of such causal relations, occurs, for example, as adding smooth roads in a landscape achieving lower resistance when moving or adding digital electronic networks so that humans can interact via social media, or changing the positioning of the earth with respect to the sun achieving hourly and seasonal differences in meteorological dynamics; e.g., (Descartes 1644; Leibniz 1698; Newton 1729; Lorenz 1963, 1993).

So, these examples illustrate that the causal relations themselves have some form of embodiment or representation within the world and if that is changing, the causal relations and their effects change accordingly. This can be described by self-modeling networks. A causal relation is by itself an abstract concept which is made concrete ('is reified'), by the world configuration on which it is based. This will be worked out for the context of (temporal-causal) network models that model such causal relationships.

8.3.2 Self-modeling Network Modeling

The concept of self-modeling network has been introduced in (Treur 2018, 2020a) and one interesting type of application is modeling adaptive networks, as shown in (Treur 2020b).

Distinction between network characteristics and network states

The following is a crucial distinction for network models:

- Network *characteristics* (such as connection weights and) have values (their strengths) and determine (e.g., mental) processes and behaviour in an implicit, automatic manner. They can be considered to provide an *embodiment view* on the network. In principle, these characteristics by themselves may not be directly accessible nor observable for network states (or a person: usually you don't see or feel a specific connection in your brain).
- Network *states* (such as sensor states, sensory representation states, preparation states, emotion states) have values (their activation levels) and are explicit representations that may be accessible for network states or a person and can be handled or manipulated explicitly. They can be considered to provide an *informational view* on the network; usually the states are assumed to have a certain informational content. In principle, for the case of a mental network, states may be accessible or observable for a person: you may see (mental image), feel (emotion) or note in some other way a specific state in your brain.

The type of network used here has been called a *temporal-causal network*, as introduced in (Treur 2016a, b). Such a network represents states X and connections $X \rightarrow Y$ between them for (causal) impacts; here states X have values $X(t)$ that usually change over time t. More precisely, the following notions form the defining *network characteristics* of a temporal-causal network model (in the form of mathematical relations and functions):

(a) **Connectivity of the network**

- *connection weights* $\boldsymbol{\omega}_{X,Y} \in [-1, 1]$ for each connection from a state X to a state Y

(b)

Aggregation of multiple impacts on a given state in the network

- *basic combination functions* $\mathbf{c}_j(..)$, $j = 1, ..., m$ for aggregation, selected for the whole network model from an available combination function library; this is done by specifying $\mathbf{mcf} = [k_1, ..., k_m]$, where k_j refers to the number combination function $\mathbf{c}_j(..)$ has within the library.
- for each state Y *combination function weights* $\boldsymbol{\gamma}_{j,Y}$ for the basic combination functions $\mathbf{c}_j(..)$, $j = 1, ..., m$ to indicate by a weighted average of the functions $\mathbf{c}_j(..)$, $j = 1, ..., m$, the aggregation of incoming single causal impacts $\boldsymbol{\omega}_{X_i,Y} X_i(t)$ of the states $X_1, ..., X_k$ from which Y gets incoming connections
- for each state Y and combination function $\mathbf{c}_j(..)$ *combination function parameters* $\boldsymbol{\pi}_{i,j,Y}$

(c)

Timing in the network

- for each state Y a *speed factor* $\boldsymbol{\eta}_Y \geq 0$.

The above defined characteristics $\boldsymbol{\omega}_{X,Y}, \boldsymbol{\gamma}_{i,Y}, \boldsymbol{\pi}_{i,j,Y}, \boldsymbol{\eta}_Y$ define in a canonical manner an associated numerical representation of the network model (Treur 2016b), Ch. 2, in difference or differential equation format which can be used for simulation and mathematical analysis:

$$Y(t + \Delta t) = Y(t) + \boldsymbol{\eta}_Y \left[\mathbf{c}_Y \left(\boldsymbol{\omega}_{X_1,Y} X_1(t), \ldots, \boldsymbol{\omega}_{X_k,Y} X_k(t)\right) - Y(t)\right] \Delta t \quad (8.1)$$

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

Here the overall combination function $\mathbf{c}_Y(..)$ for state Y is the weighted average of the basic combination functions $\mathbf{c}_j(..)$ by the specified weights $\boldsymbol{\gamma}_{j,Y}$ for Y:

$$\mathbf{c}_Y(V_1, \ldots, V_k) = \frac{\boldsymbol{\gamma}_{1,Y} \mathrm{c}_1(V_1, \ldots, V_k) + \cdots + \boldsymbol{\gamma}_{m,Y} \mathrm{c}_m(V_1, \ldots, V_k)}{\boldsymbol{\gamma}_{1,Y} + \cdots + \boldsymbol{\gamma}_{m,Y}} \quad (8.2)$$

Such equations are hidden in the dedicated software environment; see (Treur 2020b), Ch 9. Making the parameters $\boldsymbol{\pi}_{1,j,Y}$, $\boldsymbol{\pi}_{1,j,Y}$ of the basic combination functions $\mathbf{c}_{j,Y}(..)$ explicit, this becomes

$$\mathbf{c}_Y\left(\boldsymbol{\pi}_{1,1,Y}, \boldsymbol{\pi}_{2,1,Y}, \ldots, \boldsymbol{\pi}_{1,m,Y}, \boldsymbol{\pi}_{2,m,Y}, V_1, \ldots, V_k\right) \tag{8.3}$$
$$= \frac{\boldsymbol{\gamma}_{1,Y} \mathrm{c}_1\left(\pi_{1,1,Y}, \pi_{2,1,Y}, V_1, \ldots, V_k\right) + \cdots + \boldsymbol{\gamma}_{m,Y} \mathrm{c}_m\left(\pi_{1,m,Y}, \pi_{2,m,Y}, V_1, \ldots, V_k\right)}{\boldsymbol{\gamma}_{1,Y} + \cdots + \boldsymbol{\gamma}_{m,Y}}$$

There are many different approaches possible to address the issue of aggregating multiple impacts by combination functions. Therefore, for this aggregation a combination function library with a number of basic combination functions (currently more than 35) is available, while also own-defined functions can be added. Examples of basic combination functions from this library can be found in Table 8.1.

Table. 8.1 Examples of basic combination functions from the library

Combination function	**Notation**	**Formula**	**Parameters**
Identity	**id(*V*)**	V	
Complemental identity	**compid(*V*)**	$1 - V$	
Scaled sum	**ssum$_{\boldsymbol{\lambda}}$(*V*$_1$, ..., *V*$_k$)**	$\frac{V_1+\cdots+V_k}{\lambda}$	Scaling factor **$\boldsymbol{\lambda} > 0$**
Simple logistic	**slogistic$_{\boldsymbol{\sigma},\boldsymbol{\tau}}$(*V*$_1$, ...,*V*$_k$)**	$\frac{1}{1+\mathrm{e}^{-\boldsymbol{\sigma}(V_1+\ldots+V_k-\boldsymbol{\tau})}}$	Steepness **$\boldsymbol{\sigma} > 0$** Excitability threshold **$\boldsymbol{\tau}$**
Advanced logistic	**alogistic$_{\boldsymbol{\sigma},\boldsymbol{\tau}}$(*V*$_1$, ..., *V*$_k$)**	$\left[\frac{1}{1+\mathrm{e}^{-\boldsymbol{\sigma}(V_1+\ldots+V_k-\boldsymbol{\tau})}} - \frac{1}{1+\mathrm{e}^{\boldsymbol{\sigma}\boldsymbol{\tau}}}\right](1 + \mathrm{e}^{-\boldsymbol{\sigma}\boldsymbol{\tau}})$	Steepness **$\boldsymbol{\sigma} > 0$** Excitability threshold **$\boldsymbol{\tau}$**
Scaled maximun	**smax$_{\boldsymbol{\lambda}}$(*V*$_1$,..., *V*$_k$)**	$\frac{\min(V_1,\ldots,V_k)}{\lambda}$	Scaling factor **$\boldsymbol{\lambda} > 0$**
Scaled minimum	**smin$_{\boldsymbol{\lambda}}$(*V*$_1$, ..., *V*$_k$)**	$\frac{\max(V_1,\ldots,V_k)}{\lambda}$	Scaling factor **$\boldsymbol{\lambda} > 0$**
Euclidean	**eucl$_{\boldsymbol{n},\boldsymbol{\lambda}}$(*V*$_1$, ..., *V*$_k$)**	$\sqrt[n]{\frac{V_1^{\boldsymbol{n}}+\cdots+V_k^{\boldsymbol{n}}}{\lambda}}$	Order **$\boldsymbol{n} > 0$** Scaling factor **$\boldsymbol{\lambda} > 0$**
Scaled geometric mean	**sgeomean$_{\boldsymbol{\lambda}}$(*V*$_1$, ..., *V*$_k$)**	$\sqrt[k]{\frac{V_1*\ldots*V_k}{\lambda}}$	Scaling factor **$\boldsymbol{\lambda} > 0$**

Self-modeling networks connecting network characteristics and network states

As indicated above, 'network characteristics' and 'network states' are two distinct concepts for a network. A self-modeling network is a way to relate these distinct concepts to each other in an interesting and useful way:

- A *self-model* is making the implicit network characteristics (such as connection weights and excitability thresholds) explicit by adding states for these characteristics; thus the network gets an *internal self-model* of (part of) the network itself. Such additional states are called *self-model states*.

- In this way, different self-model levels are created where characteristics from one level relate to explicit states at a next self-model level. By iteration, an arbitrary number of self-model levels can be modeled, covering second-order or higher-order effects.
- Due to the above description, by adding a self-model within a network, states at a certain level refer to characteristics of the next lower level. This *referencing relation* directly relates to the self-modeling relation: state *X* models characteristic *Y* means that state *X refers to* characteristic *Y*, also phrased as state *X represents* characteristic *Y*. As these *Y* are characteristics of the network (or of a person), the representations *X* indeed form an internal *self-model* of the network (or of the person), and the referencing is also called *self-referencing*, as the referencing concerns referring to the network (or person) itself.
- Self-modeling can be applied in relation to the *physical world* by itself, but also for mental domains. For example:
  - in and by *the world*, in the brain information about causal relations between brain states is represented or encoded in physical states for connection weights, excitability thresholds or other characteristics
  - in the *mental domain*, a person can create mental states in the form of representations of his or her own (personal) characteristics, thus forming a subjective self-model (acquired by experiences).

In a wider context, this concept of self-modeling is used in different scientific areas in which it has been shown to provide substantial advantages in expressivity and in structuring and transparency of models, in particular, within AI; e.g., (Bowen and Kowalski 1982; Galton 2006; Sterling and Beer 1989; Weyhrauch 1980). Specific cases of it from a linguistic or logical perspective are representing relations between objects by objects themselves, or representing more complex statements about objects or numbers, themselves by objects or numbers, for example, like Gödel used natural numbers to represent logical statements to prove his famous incompleteness theorems for mathematical logic; e.g., (Nagel and Newman 1965; Smorynski 1977).

For network modeling, this notion turns out useful, not only to describe adaptive networks in a suitable manner (Treur 2020b), but also to model (by mental self-models) the different levels for self-referencing, self-awareness and self-interpretation as discussed in Sect. 8.2. This will be explained in more detail in the next sections. To model adaptive networks specifically, self-modeling is applied in the way that for each state *Y* of the base network, for the adaptive ones among the network structure characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_{Y}$, additional network states $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_{Y}$ (self-model states) can be introduced:

(a) **Connectivity characteristics self-model**

- self-model states $\mathbf{W}_{X_i,Y}$ are added representing adaptive connection weights $\boldsymbol{\omega}_{X_i,Y}$

(b) **Aggregation characteristics self-model**

- self-model states $\mathbf{C}_{j,Y}$ are added representing adaptive combination function weights $\boldsymbol{\gamma}_{i,Y}$
- self-model states $\mathbf{P}_{i,j,Y}$ are added representing adaptive combination function parameters $\boldsymbol{\pi}_{i,j,Y}$

(c) **Timing characteristics self-model**

- self-model states $\mathbf{H}_Y$ are added representing adaptive speed factors $\boldsymbol{\eta}_Y$.

The notations $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ for the self-model states indicate the referencing relation with respect to the characteristics $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$: here $\mathbf{W}$ refers to $\boldsymbol{\omega}$, $\mathbf{C}$ refers to $\mathbf{c}$, $\mathbf{P}$ refers to $\boldsymbol{\pi}$, and $\mathbf{H}$ refers to $\boldsymbol{\eta}$, respectively. For the processing, these self-model states define the dynamics of state Y in a canonical manner according to Eq. (8.1) whereby $\boldsymbol{\omega}_{X,Y}$, $\boldsymbol{\gamma}_{i,Y}$, $\boldsymbol{\pi}_{i,j,Y}$, $\boldsymbol{\eta}_Y$ are replaced by the state values of $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$, respectively.

An example of a representation $\mathbf{P}_{i,j,Y}$ for combination function parameter $\boldsymbol{\pi}_{i,j,Y}$ is for the excitability threshold $\boldsymbol{\tau}_Y$ of state Y; then $\mathbf{P}_{i,j,Y}$ is usually indicated by $\mathbf{T}_Y$, where $\mathbf{T}$ refers to $\boldsymbol{\tau}$. Such self-model states $\mathbf{T}_Y$ will play a role in the case study described below, as will self-model states $\mathbf{W}_{X,Y}$, referring to connection weights $\boldsymbol{\omega}_{X,Y}$. These two types of self-model states can be used to model adaptive intrinsic neuronal excitability and adaptive connection weights as described, for example, in (Chandra and Barkai 2018); e.g.:

> Learning-related cellular changes can be divided into two general groups: modifications that occur at synapses and modifications in the intrinsic properties of the neurons. While it is commonly agreed that changes in strength of connections between neurons in the relevant networks underlie memory storage, ample evidence suggests that modifications in intrinsic neuronal properties may also account for learning related behavioral changes. Long-lasting modifications in intrinsic excitability are manifested in changes in the neuron's response to a given extrinsic current (generated by synaptic activity or applied via the recording electrode). (Chandra and Barkai 2018, p. 30)

In addition to such specific self-model states $\mathbf{W}_{X,Y}$, $\mathbf{C}_{i,Y}$, $\mathbf{P}_{i,j,Y}$, $\mathbf{H}_Y$ that directly relate to the actual values of the related characteristics as described, also other self-model states can be considered that do not have such a direct relation, but still can be considered a form of self-modeling. For example, subjective self-model states can be introduced that indicate what an invidual thinks that is the connection from one state to another one. Such types of self-referencing self-model states will be discussed as well in Sect. 8.4, in the context of the designed cognitive architecture.

8.4 The Overall Cognitive Architecture

The global structure of the overall architecture is shown in Fig. 8.1. In addition to the base level it displays at different levels Self-Referencing, Self-Awareness, and Self-Interpretation, as also discussed in Sects. 8.2.1, 8.2.2, and 8.2.3, respectively.

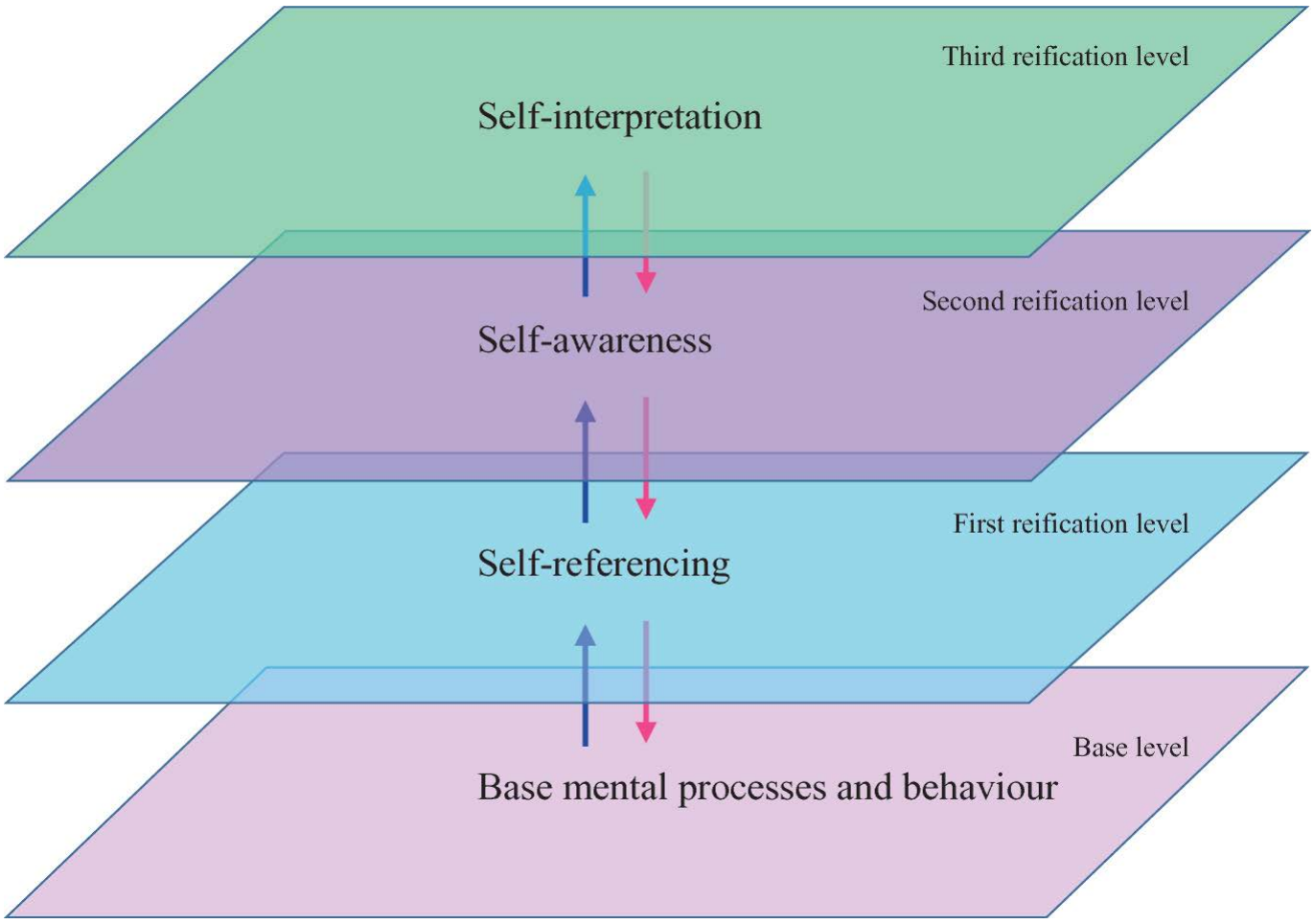


Fig. 8.1 The overall four-level cognitive architecture for self-referencing, self-awareness and self-interpretation

In some more detail this architecture can be described as follows.

8.4.1 Base Level

At the base level the mental processes of a person are considered, in the case study the client during some type of therapeutic session. These mental processes are modeled through causal connections between mental states that drive their dynamics. Using the concept of temporal-causal network for that, a number of network characteristics are specified that determine the person's mental functioning; by Eq. (8.1) these characteristics define the base level's dynamics for the mental processes. The characteristics such as connection weights and excitability thresholds may cover, for example:

- tendencies to do something or avoid something in terms of actions
- connections associating emotions to events
- the sensitivity for responding.

For the case study, the base network has the following network connections (see the case description in Sect. 8.2.5 used as a point of departure):

- $srs_{Complaints} \rightarrow ps_{GotoGP}$ complaints leads to preparation to go to GP
- $srs_{Complaints}$, $ms_{Stigma} \rightarrow bs_{Stigma}$ complaints and memory about stigma lead to the belief state that confirming a mental problem provides a stigma
- $bs_{Stigma} \rightarrow ds_{AvoidConfirmation}$ belief bs_{Stigma} leads to desire state $ds_{AvoidConfirmation}$ to avoid confirming mental problems
- $srs_{Complaints} \rightarrow bs_{Irrelevant}$ complaints lead to the belief $bs_{Irrelevant}$ that after a while mental problems go away by themselves
- $bs_{Irrelevant}$, $ds_{AvoidConfirmation} \rightarrow ps_{AvoidMentalHelp}$ the desire to avoid confirming mental problems, and belief $bs_{Irrelevant}$ make preparing to avoid mental help $ps_{AvoidMentalHelp}$

In Fig. 8.2 a graphical representation of these causal relations for the case study is depicted. For a brief explanation of all states for the case study, see Table 8.3.

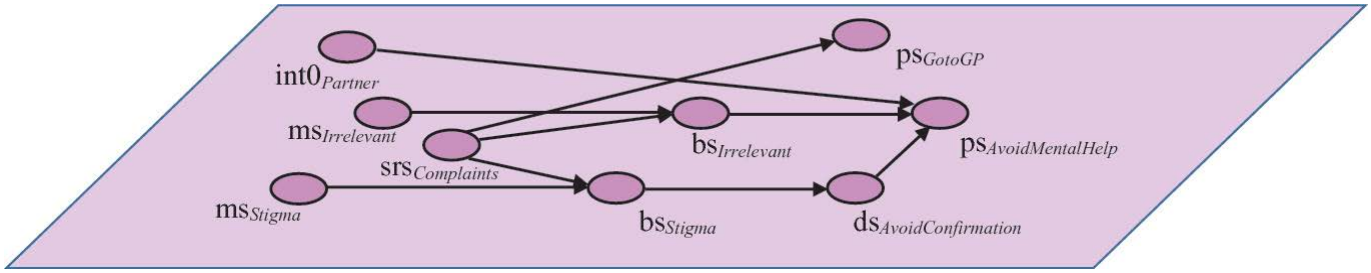


Fig. 8.2 Base level connections for the considered case study

8.4.2 First Self-model Level: Self-referencing

At the first self-model level a number of representation states for a first-order self-model of the own base level characteristics are included. These representations can be for the physical world or (subjectively) relate to the person's mental processes.

Representation states $\mathbf{W}_{X,Y}$

For the physical world, **W**-states are used as already mentioned in Sect. 8.3.2 above, where $\mathbf{W}_{X,Y}$ represents the actual connection weight for the connection from state *X* to state *Y*. As an example, $\mathbf{W}_{\mathrm{srs}_{Complaints},\mathrm{bs}_{Stigma}}$ indicates the weight for the connection from $\mathrm{srs}_{Complaints}$ to bs_{Stigma}. This explicit representation can be used for adaptation, for example, based on Hebbian learning (Hebb 1949).

Representation states $\mathbf{RW}_{X,Y}$

Related to the individual, for example, the own known (or believed) connection weight characteristics are represented by **RW**-states, where $\mathbf{RW}_{X,Y}$ indicates the person's representation for the weight of the connection from base state *X* to base state *Y*. Here the prefix **R** for representation is used to distinguish it from the actual value used in the processing of the network model: the own representations $\mathbf{RW}_{X,Y}$ may be quite different from the 'real' values $\mathbf{W}_{X,Y}$. This knowledge has to be acquired by experience, which may depend on the situations the person actually encounters. For example, for a never experienced situation a person may not know at all what own responses will be triggered by it. Examples for the case study are shown in Table 8.2. Note that all these states in Table 8.2 in principle may occur without being aware of them. Awareness states will be addressed at the next level, in Sect. 8.4.3. In the example model for the case study described below, for the sake of simplicity only some of these states are included, as the others do not play a role in the example scenario addressed.

Table. 8.2 Examples of self-representation states **RW** and self-referencing states **SRW**

Self-model state	Informal explanation
$\mathbf{RW}\mathrm{srs}_{Complaints},\mathrm{ps}_{GotoGP}$	I believe I have a tendency to go to a GP upon complaints
$\mathbf{RW}\mathrm{srs}_{Complaints},\mathrm{bs}_{Stigma}$ $\mathbf{RW}\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}$	I believe I have tendencies to believe, due to complaints and memories, respectively, that confirming a mental problem provides a stigma
$\mathbf{RW}\mathrm{bs}_{Stigma},\mathrm{ds}_{AvoidConfirmation}$	I believe I have a tendency to desire, due to believing that confirming a mental problem provides a stigma, to avoid confirming mental problems
$\mathbf{RW}\mathrm{srs}_{Complaints},\mathrm{bs}_{Irrelevant}$	I believe I have a tendency to believe upon complaints that mental problems often go away by themselves after a while

Self-model state	Informal explanation
RWds$_{AvoidConfirmation}$,ps$_{AvoidMentalHelp}$ **RW**bs$_{Irrelevant}$,ps$_{AvoidMentalHelp}$	I believe I have tendencies to avoid mental help due to a desire to avoid confirming mental problems, and believing that after a while mental problems usually go away by themselves, respectively
SRWsrs$_{Complaints}$,bs$_{Stigma}$	I believe that I believe that confirming a mental problem will provide a stigma because I have a tendency to believe that upon complaints
SRWms$_{Stigma}$,bs$_{Stigma}$	I believe that I believe that confirming a mental problem will provide a stigma because I have a tendency to believe that due to similar memories from past situations
SRWms$_{Irrelevant}$,bs$_{Irrelevant}$	I believe that I believe that mental problems will go away by themselves because I have a tendency to believe that due to similar memories from past situations

Table. 8.3 States used in the computational case study and their explanation

State nr	State name	Explanation	Level
X_1	$int0_{Partner}$	Intervention 0 by the partner to encourage not to avoid mental help	Base level
X_2	$srs_{Complaints}$	Sensory representation state for complaints	
X_3	ms_{Stigma}	Memory state about someone getting a stigma due to confirming mental problems	
X_4	$ms_{Irrelevant}$	Memory state about someone considering similar mental problems irrelevant	
X_5	bs_{Stigma}	Belief state for getting a stigma if confirming mental problems	
X_6	$bs_{Irrelevant}$	Belief state that mental problems go away by themselves	
X_7	$ds_{AvoidConfirmation}$	Desire state to avoid confirmation of mental problems	
X_8	ps_{GotoGP}	Preparation state for going to the GP	
X_9	$ps_{AvoidMentalHelp}$	Preparation state for avoiding mental help	
X_{10}	$\mathbf{W}ms_{Stigma}, bs_{Stigma}$	Self-model state for the weight of the connection from ms_{Stigma} to bs_{Stigma}	First-order self-model level: Self-model and Self-referencing
X_{11}	$\mathbf{W}ms_{Irrelevant}, bs_{Irrelevant}$	Self-model state for the weight of the connection from $ms_{Irrelevant}$ to $bs_{Irrelevant}$	
X_{12}	$\mathbf{RW}srs_{Complaints}, bs_{Stigma}$	Self-representation state for the person's belief on the weight of the connection from $srs_{Complaints}$ to bs_{Stigma}	
X_{13}	$\mathbf{RW}ms_{Stigma}, bs_{Stigma}$	Self-representation state for the person's belief on the weight of the connection from ms_{Stigma} to bs_{Stigma}	
X_{14}	$\mathbf{RW}ms_{Irrelevant}, bs_{Irrelevant}$	Self-representation state for the person's belief on the weight of the connection from $ms_{Irrelevant}$ to $bs_{Irrelevant}$	
X_{15}	$\mathbf{SRW}srs_{Complaints}, bs_{Stigma}$	Self-referencing state for the state bs_{Stigma} and the weight of the connection from $srs_{Complaints}$ to bs_{Stigma}	
X_{16}	$\mathbf{SRW}ms_{Stigma}, bs_{Stigma}$	Self-referencing state for the state bs_{Stigma} and the weight of the connection from ms_{Stigma} to bs_{Stigma}	
X_{17}	$\mathbf{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	Self-referencing state for the state $bs_{Irrelevant}$ and the weight of the connection from $ms_{Irrelevant}$ to $bs_{Irrelevant}$	
X_{18}	$int1_{Therapist}$	Intervention 1 by the therapist to get awareness of the role of own memories and of comforting ideas	Second-order self-model level: Self-awareness
X_{19}	$\mathbf{F}_{\mathbf{SRW}}x, bs_{Stigma}$	Focus for self-referencing for the generation of the belief state bs_{Stigma}	
X_{20}	$\mathbf{A}Comfortingideas$	Awareness state for comforting ideas	
X_{21}	$\mathbf{SA}_{\mathbf{SRW}}srs_{Complaints}, bs_{Stigma}$	Awareness state of how the belief state bs_{Stigma} relates to an own characteristic connecting to the complaints	
X_{22}	$\mathbf{SA}_{\mathbf{SRW}}ms_{Stigma}, bs_{Stigma}$	Awareness state of how the belief state bs_{Stigma} relates to a memory state concerning stigma	
X_{23}	$\mathbf{SA}_{\mathbf{SRW}}ms_{Irrelevant}, bs_{Irrelevant}$	Awareness state of how the belief state $bs_{Irrelevant}$ relates to a memory state concerning irrelevance of certain mental problems	
X_{24}	$\mathbf{T}_{\mathbf{SRW}}srs_{Complaints}, bs_{Stigma}$	Self-model state for modulating activation of the state $\mathbf{SRW}srs_{Complaints}, bs_{Stigma}$ via its excitability threshold	
X_{25}	$\mathbf{T}_{\mathbf{SRW}}ms_{Stigma}, bs_{Stigma}$	Self-model state for modulating activation of the state $\mathbf{SRW}ms_{Stigma}, bs_{Stigma}$ via its excitability threshold	
X_{26}	$\mathbf{T}_{\mathbf{SRW}}ms_{Irrelevant}, bs_{Irrelevant}$	Self-model state for modulating activation of the state $\mathbf{SRW}ms_{Irrelevant}, bs_{Irrelevant}$ via its excitability threshold	
X_{27}	$int2_{Therapist}$	Intervention 2 by the therapist to focus more on getting awareness of the role of own memories for belief state bs_{Stigma}	Third-order self-model level: Self-interpretation
X_{28}	$\mathbf{F}_{\mathbf{SASRW}}x, bs_{Stigma}$	Focus state for intervention 2	
X_{29}	$\mathbf{T}_{\mathbf{SASRW}}ms_{Stigma}, bs_{Stigma}$	State for modulating awareness about the role of the own memories by a decreased excitability threshold	
X_{30}	$\mathbf{V}^{-}_{\mathbf{SASRW}}ms_{Stigma}, bs_{Stigma}$	Negative valuation state for the awareness of the role of the own memory connections for the belief state bs_{Stigma}	
X_{31}	$\mathbf{V}^{+}\mathbf{A}Comfortingideas$	Positive valuation state for awareness of comforting ideas	
X_{32}	$\mathbf{V}^{-}\mathbf{W}x, bs_{Stigma}$	Negative valuation of the own connections generating the belief state bs_{Stigma}	
X_{33}	$\mathbf{T}bs_{Stigma}$	Self-model state for modulating activation of the belief state bs_{Stigma} by increasing its excitability threshold	

In Fig. 8.3, the self-model states modeling the first-order self-model by **RW**-states are depicted together with the representation (or referencing) relations to the connections they represent (dashed yellow lines). Here the base level network characteristic for the weight $\boldsymbol{\omega}_{X,Y}$ of a connection from a base state X to a base state Y is represented by a first self-model level state denoted by $\mathbf{RW}_{X,Y}$.

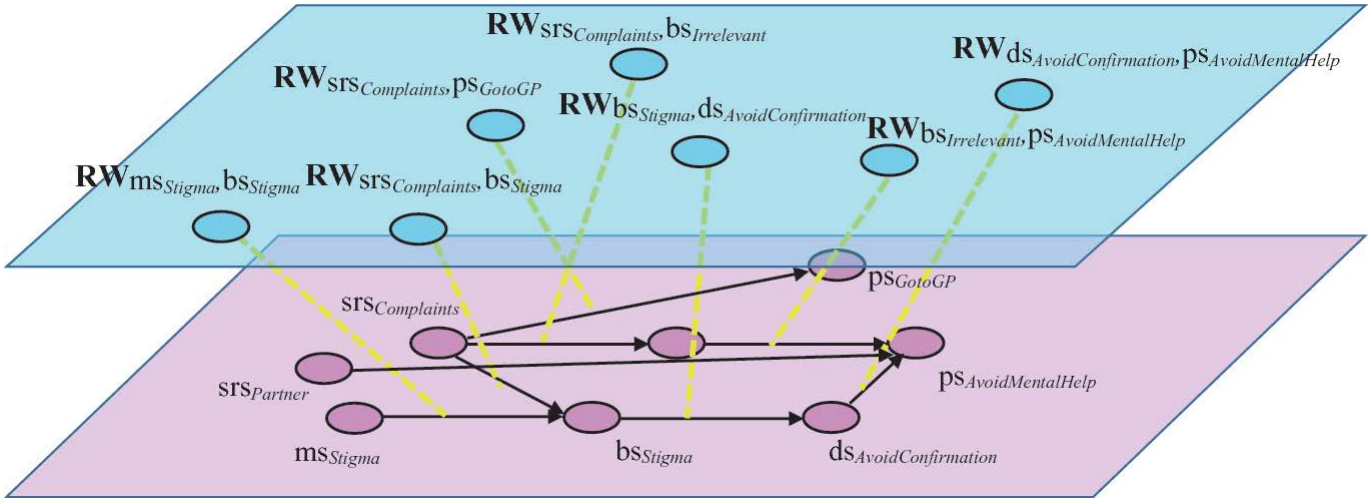


Fig. 8.3 First-order self-model level states for the example and their representation or referencing relations

Note that these referencing relations are not used in the processing of the model. Instead, upward causal connections from base states to first-order self-model states representing their connections are used as depicted in Fig. 8.4. Based on these connections, the first-order self-referencing states can be learnt from experiences and in that way get their values. For non-accessible or badly accessible (blind spots) base states, these upward connections are weak or very weak, or even nonexistent.

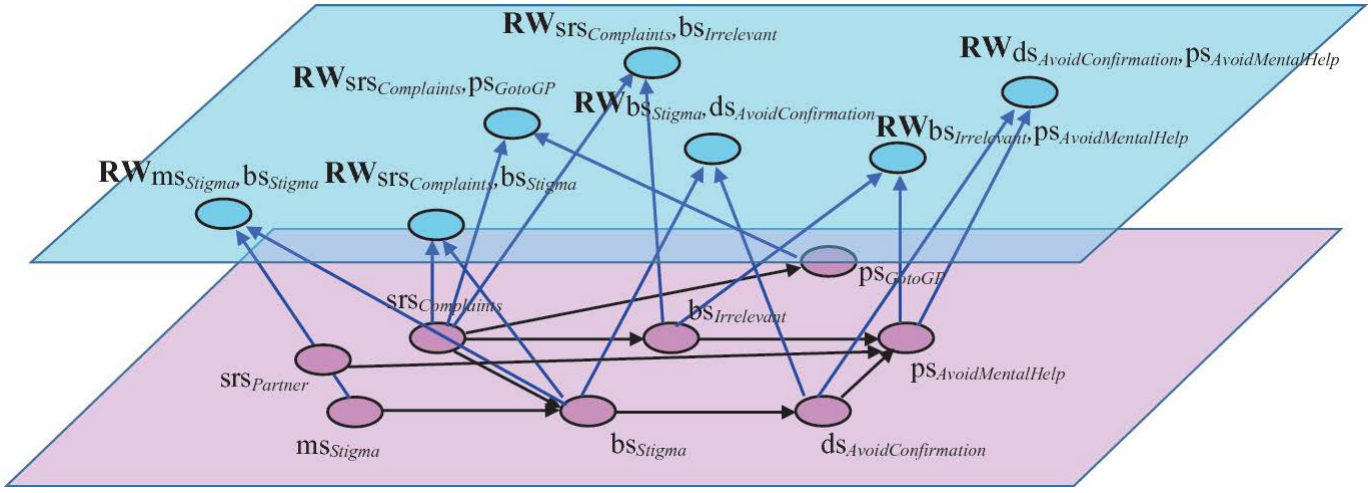


Fig. 8.4 Base level and first-order self-model level: upward interlevel connections for the example, assuming full accessibility of the base states (but no accessibility of connections)

Self-referencing states SRW $_{X,Y}$

Given an actually occurring mental state *Y*, this, together with the relevant characteristics representation $\mathbf{RW}_{X,Y}$, may trigger *first-order self-referencing states* $\mathbf{SRW}_{X,Y}$ which indicate the person knowing or believing that mental state *Y* occurs and that the occurrence of this mental state *Y* relates to certain personal characteristics of him or herself. The first-order self-referencing states considered here are

> $\mathbf{SRW}_{X_1,Y}$, ..., $\mathbf{SRW}_{X_k,Y}$ for personal characteristics in the form of *weights of the incoming connections* of state *Y* from states X_1, .., X_k, respectively.

To generate these self-referencing states, causal connections are applied to $\mathbf{H}_{\mathbf{WT}_{\text{exhrisk}}}$ both from the considered base state *Y* and from the relevant characteristic $\mathbf{RW}_{X_i,Y}$:

$$Y \rightarrow \mathbf{H}_{\mathbf{WT}_{\text{exhrisk}}}$$
$$\mathbf{RW}_{X_i,Y} \rightarrow \mathbf{H}_{\mathbf{WT}_{\text{exhrisk}}}$$

At the first self-model level, from the base level network characteristics not only connection weights can be explicitly represented by self-model states, but also other characteristics of the base level network, such as the excitability threshold $\boldsymbol{\tau}_X$ of a base state *X* that plays an important role in sensitivity of responses of *X*. For example, the sensitivity threshold $\boldsymbol{\tau}_{\text{bs}_{Stigma}}$ of base state bs_{Stigma} can be represented by a representation state $\mathbf{RT}_{\text{bs}_{Stigma}}$ and self-referencing state $\mathbf{SRT}_{\text{bs}_{Stigma}}$ at the first self-model level. For the moment, in this exploration we limit ourselves to self-modeling of the connection weights.

8.4.3 Second Self-model Level: Self-awareness

An awareness state of a state *Y* refers to *Y* and therefore can be modeled at one self-model level higher than *Y*. An awareness state for state *Y* can lead to a change of network characteristics related to *Y*, for example, its excitability threshold or weights of incoming connections of *Y*. As awareness states are modeled at one level higher than *Y*, the effect that awareness strengthens learning can be modeled by causal relations within that level. Via downward causal connections, this can lead to a decreased excitability threshold or increased weights of incoming connections used for *Y*. Note that, if *Y* is mental state at the base level, awareness of *Y* by itself is not considered self-awareness as there is no self-referencing. However, awareness of a self-referencing state for *Y*, such as

$\mathbf{SRW}_{X_i,Y}$ indicating that Y occurs (partly) because X_i has an effect on Y, is modeled by a self-awareness state $\mathbf{SA}_{\mathbf{SRW}_{X_i,Y}}$. As it refers to a level 1 state $\mathbf{SRW}_{X_i,Y}$, this self-awareness state $\mathbf{SA}_{\mathbf{SRW}_{X_i,Y}}$ is inherently second-order. As a particular case, this can be applied to an emotion state Y, in which case $\mathbf{SA}_{\mathbf{SRW}_{X_i,Y}}$ indicates that the person not only has unconscious self-knowledge that state X_i contributes to triggering of emotion Y, but is also aware that (s)he has the characteristic or tendency that state X_i contributes to the triggering of emotion Y, which indeed is a form of self-awareness.

Generation of any awareness state for a given state Y at any level is modeled in a practical manner by applying a combination of three principles occurring in multiple consciousness theories:

- focusing of attention; e.g., (Graziano 2013; Graziano et al. 2019)
- a winner-takes-it-all competition between states; e.g., (Minsky 1986; Baars 1997; Graziano et al. 2019)
- enhanced accessibility due to awareness; e.g., (Minsky 1986; Baars 1997; Graziano et al. 2019).

For the latter, an awareness state for Y may amplify the activation of Y via changes of some of the network characteristics related to Y, such as its incoming causal connections, or excitability threshold. By such mechanisms, the second-order self-model states may also affect the connections for the first-order self-model states by which knowledge on self-referentiality is obtained.

8.4.4 Third Self-model Level: Self-interpretation

Self-interpretation is considered here as analysis of the own processes in relation to the person's conscious self-model. This conscious self-model is represented by the self-awareness states at the second self-model level. Such an analysis refers to these self-awareness states and is therefore represented at the third self-model level. Some of the states at this third self-model level are positive and negative valuation states $\mathbf{V}^+_Y$ and $\mathbf{V}^-_Y$ where Y is a second-order state. For example, a state Y with high negative valuation can be considered undesirable. This enables the possibility that due to the self-interpretation analysis, some changes in the characteristics 'embodying' the second-order self-model are made. For example, incoming connections of some of the awareness states might get changed weights or the excitability thresholds of some of them might change due to the self-interpretation. But also characteristics of the lower levels (first self-model level, base level) may be affected in a similar manner. For example, the self-interpretation may reveal that there is an undesirable lack of sensitivity of a certain state Y, and then via some form of remedy that state Y is

made more sensitive, modeled by decrease of its excitability threshold. Or, the other way around, a base state analysed as undesirable is blocked or suppressed by increasing its excitability threshold. Or, alternatively, a suppressing pathway to this state is strengthened, like it can happen in emotion regulation. Depending on how detailed such a remedy is modeled, its effect can either be modeled as a direct influence from the third to the first self-model level representing the characteristics, or via intermediate processes also involving the second self-model level: first an effect from the third on the second self-model level, which in turn has an effect on the first self-model level. More specific examples of this will be shown in Sect. 8.5.

8.5 The Four-Level Self-modeling Network Model for the Case Study

In this section, the self-modeling network model for the cognitive architecture from Fig. 8.1 and the obtained computational model for the case study discussed in Sect. 8.2 is presented in some more detail. For the sake of understanding, the multilevel network model will be presented level by level and the connectivity will be shown in Figs. 8.5, 8.6 and 8.7 accordingly. In Table 8.3 an overview of the states used and their explanation can be found.

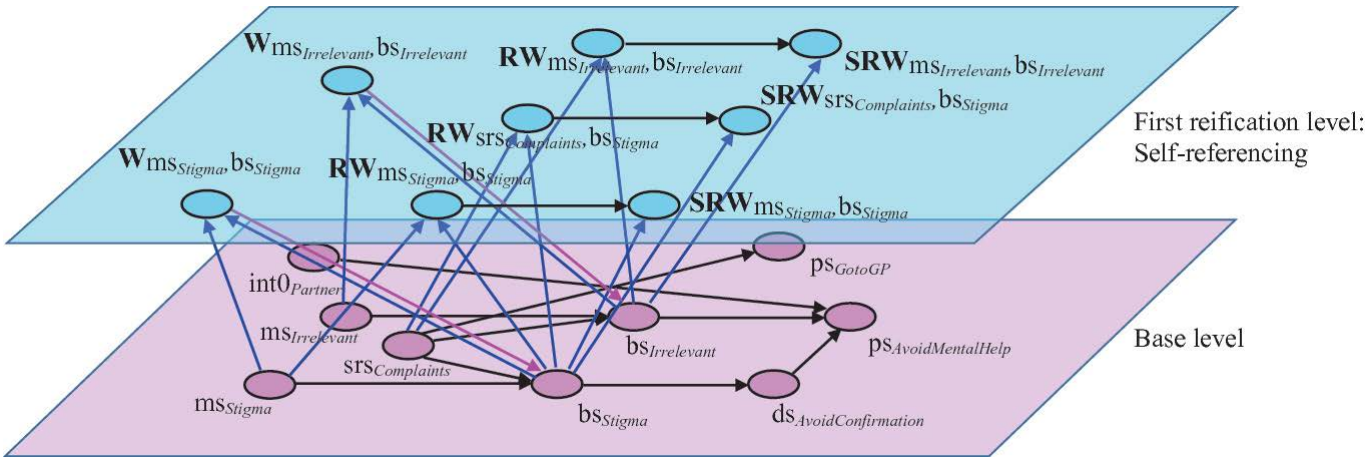


Fig. 8.5 Connectivity of the base level and first self-model level (self-referencing) for the example

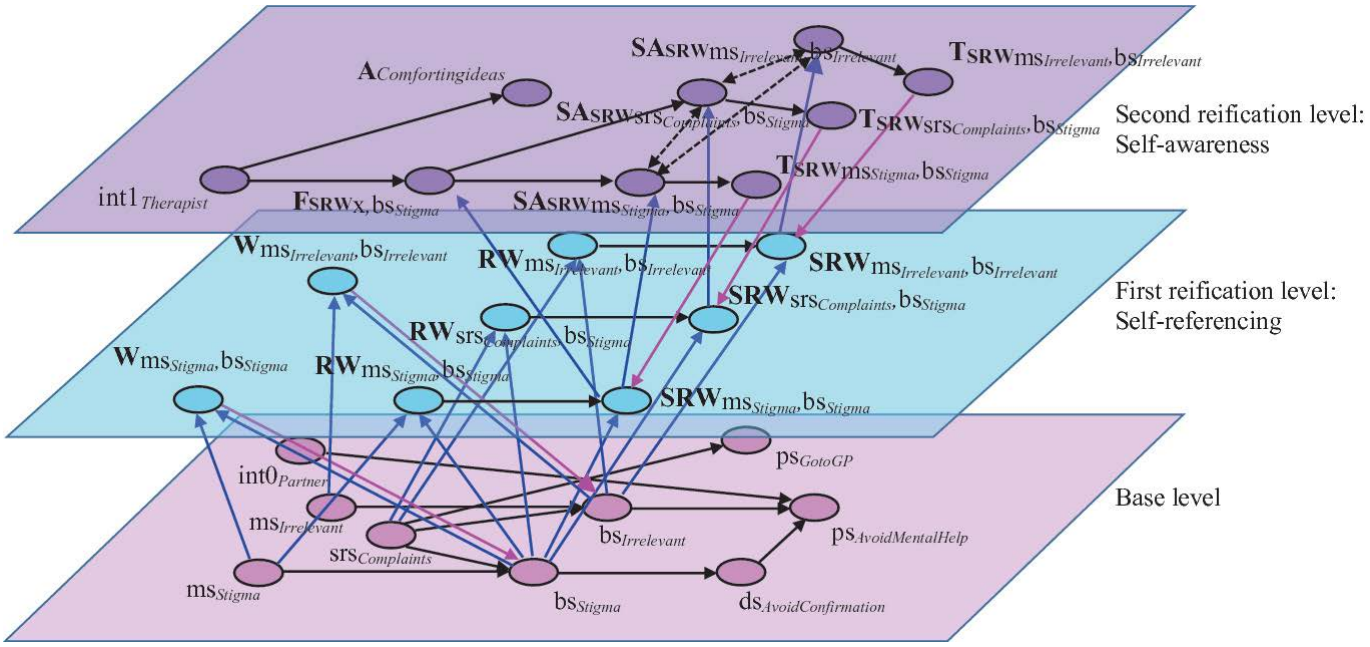


Fig. 8.6 Connectivity of the base level and first (self-referencing) and second (self-awareness) self-model level for the example

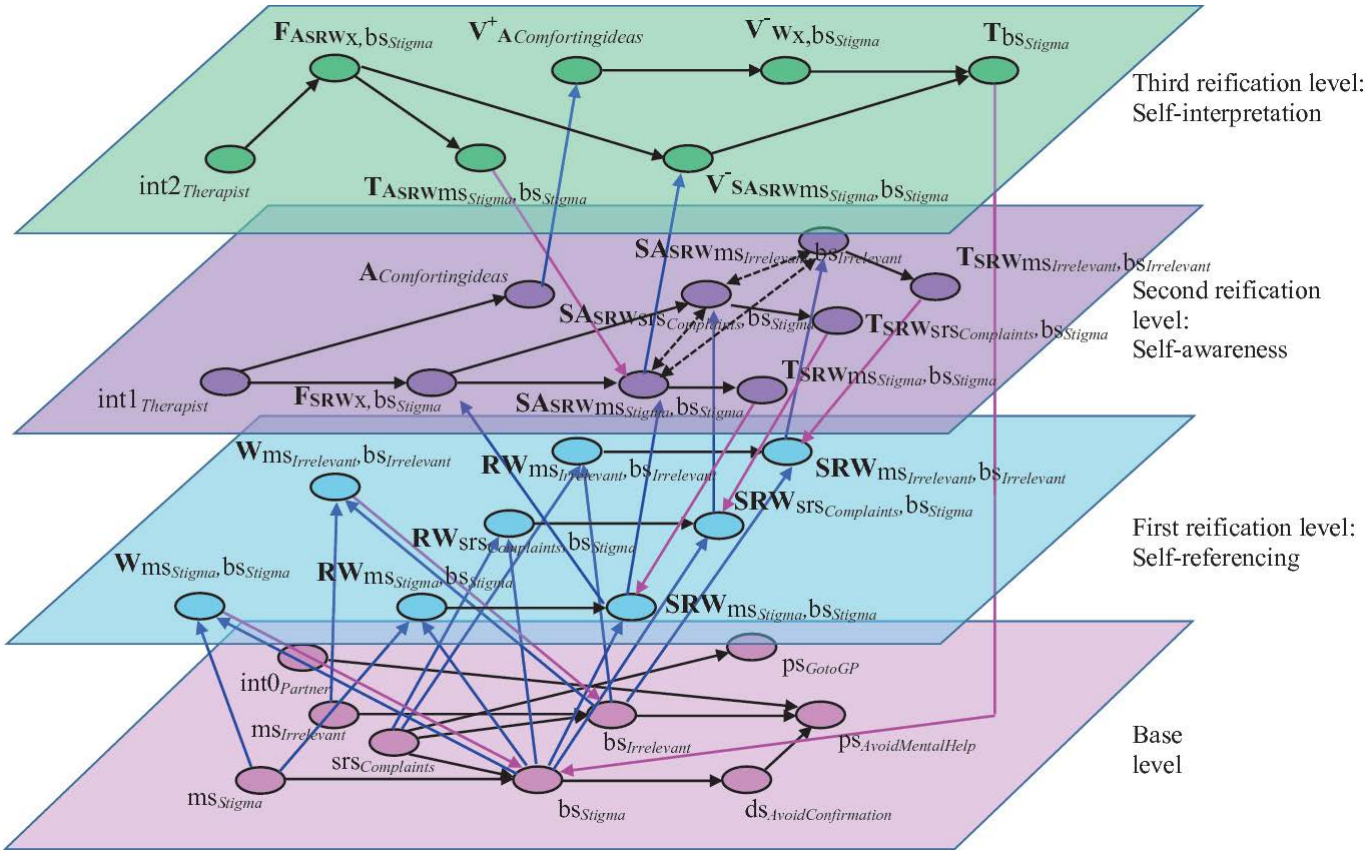


Fig. 8.7 Connectivity of the overall network model. Base level and first (self-referencing states), second (self-awarenss states) and third (self-interpretation states) self-model level

Base level network states

The base network is as displayed in Fig. 8.2 above and discussed in Sect. 8.4.1. This base network represents the client's mental processes. Note that also input

from the client's partner is incorporated, indicated by $\text{int0}_{Partner}$. This models that the partner tries to persuade the client to go for mental help; it has a suppressing effect on $\text{ps}_{AvoidMentalHelp}$. The base network is depicted in Fig. 8.5 as the lower plane.

First self-model level network states: first-order representation and self-referencing

For the first self-model level **RW**-states are used as representations for the own characteristics, in this case in the form of the connections to the two belief states bs_{Stigma} and $\text{bs}_{Irrelevant}$; see Fig. 8.5. Based on these **RW**-states, self-referencing **SRW**-states are generated that describe that the respective belief state occurs due to the respective personal characteristic by a connection (recall that $\mathbf{SRW}_{X,Y}$ stands for a self-referencing state concerning state Y and the connection from X to Y).

Second self-model level network states: self-awareness

Part of the self-referencing states $\mathbf{SA}_{\mathbf{SRW}_{X_i,Y}}\mathbf{SRW}_{X,Y}$ activate awareness states $\mathbf{SA}_{\mathbf{SRW}_{X,Y}}$ for them at the second self-model level, depending on attentional focusing and a winner-takes-it-all-competition; see Fig. 8.6. Here the therapist initiates intervention $\text{int1}_{Therapist}$, which makes the person focus (via focus state $\mathbf{F}_{\mathbf{SRW}_{\text{x,bs}_{Stigma}}}$) on the belief state bs_{Stigma} about stigma and possible relevant memories the person has for it. This has no substantial effect yet. Also, part of the therapist's intervention $\text{int1}_{Therapist}$ is making the person aware of some comforting ideas, for example, that nowadays people are happy to openly talk about their mental problems and are appreciated for that instead that it harms them. This intervention 1 is aiming at two awareness states to become active that were not active before the intervention: $\mathbf{A}_{Comfortingideas}$ and the awareness state about having a memory connection $\mathbf{SA}_{\mathbf{SRW}_{\text{ms}_{Stigma},\text{bs}_{Stigma}}}$. Only the first one actually becomes active.

Third self-model level network states: self-interpretation

Next, self-interpretation is modeled at the third self-model level; see Fig. 8.7. Here intervention $\text{int2}_{Therapist}$ is initiated by the therapist by suggesting to take time to focus and concentrate more (perhaps using some specific techniques) on becoming aware of memories from the past about a stigma (state $\mathbf{F}_{\mathbf{SA}_{\mathbf{SRW}_{\text{x,bs}_{Stigma}}}}$).

This leads to the person opening up more for such an awareness by decreasing the excitability threshold for it via state $\mathbf{T}_{\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}}$ and the pink downward arrow from it. This has effect and the person becomes aware of a memory that triggers his belief about a negative stigma; so it is only now that self-awareness state $\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}$ becomes active.

Note that this state represents that the person is aware that his belief about a stigma for mental problems has as one of its triggers a connection (a personal characteristic) from a bad memory from long ago when the attitude toward mental problems in general was very different from nowadays. Another process at the third self-model level for interpretation is valuation of the various activated awareness states. This leads to negative valuation state (a $\mathbf{V}^{-}$-state) of the belief state for stigma and a positive valuation state (a $\mathbf{V}^{+}$-state) for the awareness of the comforting ideas. These valuations lead to a state $\mathrm{T}_{\mathrm{bs}_{Stigma}}$ meant to suppress this belief state bs_{Stigma} by increasing its excitability threshold. By a downward connection (the long pink arrow), this suppression is actually performed, so that now this belief bs_{Stigma} is not generated anymore.

8.6 Detailed Specification

In this section the full details of the self-modeling network model for the cognitive architecture from Fig. 8.1 are discussed. In Sect. 8.4 the *connectivity characteristics* of the network were shown in a graphical manner, but, for example, no connection weights, nor details about the *aggregation characteristics* and *timing characteristics* were shown. Here all three types of characteristics will be shown in full detail in *role matrix format*, which makes the presented work reproducible; see Boxs 8.1, 8.2, and 8.3.

	mb base connectivity	1	2	3	4
X_1	$int0_{Partner}$	X_1			
X_2	$srs_{Complaints}$	X_2			
X_3	ms_{Stigma}	X_3			
X_4	$ms_{Irrelevant}$	X_4			
X_5	bs_{Stigma}	X_2	X_3		
X_6	$bs_{Irrelevant}$	X_4			
X_7	$ds_{AvoidConfirmation}$	X_5			
X_8	ps_{GotoGP}	X_2			
X_9	$ps_{AvoidMentalHelp}$	X_6	X_7	X_1	
X_{10}	$\mathbf{W}ms_{Stigma}, bs_{Stigma}$	X_3	X_5	X_{10}	
X_{11}	$\mathbf{W}ms_{Irrelevant}, bs_{Irrelevant}$	X_4	X_6	X_{11}	
X_{12}	$\mathbf{RW}srs_{Complaints}, bs_{Stigma}$	X_2	X_5	X_{12}	
X_{13}	$\mathbf{RW}ms_{Stigma}, bs_{Stigma}$	X_3	X_5	X_{13}	
X_{14}	$\mathbf{RW}ms_{Irrelevant}, bs_{Irrelevant}$	X_4	X_6	X_{14}	
X_{15}	$\mathbf{SRW}srs_{Complaints}, bs_{Stigma}$	X_5	X_{12}		
X_{16}	$\mathbf{SRW}ms_{Stigma}, bs_{Stigma}$	X_5	X_{13}		
X_{17}	$\mathbf{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	X_6	X_{14}		
X_{18}	$int1_{Therapist}$	X_{18}			
X_{19}	$\mathbf{F}_{SRW}x, bs_{Stigma}$	X_{16}	X_{18}		
X_{20}	$\mathbf{A}_{Comfortingideas}$	X_{18}			
X_{21}	$\mathbf{SA}_{SRW}srs_{Complaints}, bs_{Stigma}$	X_{15}	X_{22}	X_{23}	
X_{22}	$\mathbf{SA}_{SRW}ms_{Stigma}, bs_{Stigma}$	X_{16}	X_{21}	X_{23}	X_{18}
X_{23}	$\mathbf{SA}_{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	X_{17}	X_{21}	X_{22}	
X_{24}	$\mathbf{T}_{SRW}srs_{Complaints}, bs_{Stigma}$	X_{21}			
X_{25}	$\mathbf{T}_{SRW}ms_{Stigma}, bs_{Stigma}$	X_{22}			
X_{26}	$\mathbf{T}_{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	X_{23}			
X_{27}	$int2_{Therapist}$	X_{27}			
X_{28}	$\mathbf{F}_{SASRW}x, bs_{Stigma}$	X_{27}			
X_{29}	$\mathbf{T}_{SASRW}ms_{Stigma}, bs_{Stigma}$	X_{28}			
X_{30}	$\mathbf{V}^{-}_{SASRW}ms_{Stigma}, bs_{Stigma}$	X_{22}	X_{28}		
X_{31}	$\mathbf{V}^{+}\mathbf{A}_{Comfortingideas}$	X_{20}			
X_{32}	$\mathbf{V}^{-}wx, bs_{Stigma}$	X_{31}			
X_{33}	$\mathbf{T}bs_{Stigma}$	X_{30}	X_{32}		

	mcw connection weights	1	2	3	4
X_1	$int0_{Partner}$	1			
X_2	$srs_{Complaints}$	1			
X_3	ms_{Stigma}	1			
X_4	$ms_{Irrelevant}$	1			
X_5	bs_{Stigma}	0.1	X_{10}		
X_6	$bs_{Irrelevant}$	X_{11}			
X_7	$ds_{AvoidConfirmation}$	1			
X_8	ps_{GotoGP}	1			
X_9	$ps_{AvoidMentalHelp}$	1	1	-0.5	
X_{10}	$\mathbf{W}ms_{Stigma}, bs_{Stigma}$	1	1	1	
X_{11}	$\mathbf{W}ms_{Irrelevant}, bs_{Irrelevant}$	1	1	1	
X_{12}	$\mathbf{RW}srs_{Complaints}, bs_{Stigma}$	1	1	1	
X_{13}	$\mathbf{RW}ms_{Stigma}, bs_{Stigma}$	1	1	1	
X_{14}	$\mathbf{RW}ms_{Irrelevant}, bs_{Irrelevant}$	1	1	1	
X_{15}	$\mathbf{SRW}srs_{Complaints}, bs_{Stigma}$	1	1		
X_{16}	$\mathbf{SRW}ms_{Stigma}, bs_{Stigma}$	1	1		
X_{17}	$\mathbf{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	1	1		
X_{18}	$int1_{Therapist}$	1			
X_{19}	$\mathbf{F}_{SRW}x, bs_{Stigma}$	1	1		
X_{20}	$\mathbf{A}_{Comfortingideas}$	1			
X_{21}	$\mathbf{SA}_{SRW}srs_{Complaints}, bs_{Stigma}$	0.7	-0.5	-0.5	
X_{22}	$\mathbf{SA}_{SRW}ms_{Stigma}, bs_{Stigma}$	0.7	-0.5	-0.5	0.5
X_{23}	$\mathbf{SA}_{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	0.8	-0.5	-0.5	
X_{24}	$\mathbf{T}_{SRW}srs_{Complaints}, bs_{Stigma}$	0.8			
X_{25}	$\mathbf{T}_{SRW}ms_{Stigma}, bs_{Stigma}$	0.8			
X_{26}	$\mathbf{T}_{SRW}ms_{Irrelevant}, bs_{Irrelevant}$	0.8			
X_{27}	$int2_{Therapist}$	1			
X_{28}	$\mathbf{F}_{SASRW}x, bs_{Stigma}$	1			
X_{29}	$\mathbf{T}_{SASRW}ms_{Stigma}, bs_{Stigma}$	1			
X_{30}	$\mathbf{V}^{-}_{SASRW}ms_{Stigma}, bs_{Stigma}$	1	1		
X_{31}	$\mathbf{V}^{+}\mathbf{A}_{Comfortingideas}$	1			
X_{32}	$\mathbf{V}^{-}wx, bs_{Stigma}$	1			
X_{33}	$\mathbf{T}bs_{Stigma}$	1	1		

Box 8.1 Role matrices **mb** and **mcw** for the connectivity characteristics of the adaptive network model

Each role matrix addresses a specific role of the specified values of the characteristics. In each role matrix, each state X_i in the network has its own row. In this row in the different columns for the specific state X_i the different causal impacts are listed from this role. This format can be used as input for the dedicated software environment that is available and enables simulation. First, it is shown how the graphical representation of Fig. 8.7 can be expressed in role matrix **mb** for base connectivity. Next, to add more detail, in role matrix **mcw** numbers for the connection weights are specified. Both these role matrices **mb**

and **mcw** are shown in Box 8.1 and fully specify all connectivity characteristics. Role matrix **mb** specifies the causal impacts from a *base role*; these are the other states that have a basic impact on the state X_i addressed in that row. For example, state X_5 (= bs_{Stigma}), which is the belief state bs_{Stigma}, gets basic causal impact from the complaints representation state X_2 (= $srs_{Complaints}$) in the first column and the memory state X_3 (= ms_{Stigma}) in the second column. In matrix **mcw**, from the connection weight role, two more causal impacts on state X_3 (= bs_{Stigma}) are specified, namely the causal impact from the connection weight role for X_2 (= $srs_{Complaints}$) specified by the 1 in the first column of **mcw** and for X_3 (= ms_{Stigma}) specified by the 1 in the second column of **mcw**. So, these numbers specify the basic connection weights used for state X_5 (= bs_{Stigma}). In this case, most of these causal impacts are constant, but the connection weights from the memory states to the belief states were made adaptive by adding self-model states X_{10} (= $\mathbf{W}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}$) and X_{11} (= $\mathbf{W}_{\mathrm{ms}_{Irrelevant},\mathrm{bs}_{Irrelevant}}$) for them, so that these weight numbers and their causal impact from that role exerted on X_5 (= bs_{Stigma}) and X_6 (= $bs_{Irrelevant}$) changes over time (modeled based on Hebbian learning). Note that the values −0.5 in **mcw** for the mutual connections for X_{21} to X_{23} model the competition between these three self-awareness states.

Note that in role matrix **mb** only the horizontal, intralevel connections (black arrows) and upward interlevel connections (blue arrows) from Fig. 8.7 are specified, as only these indicate basic impact. The downward interlevel connections (pink arrows) indicate a special effect according to a different, non-basic role: the role played by the source state, which is specified in the role matrix for that specific role. Below, this will be discussed for the **T**-states X_{33} (= $\mathbf{H}_{\mathrm{W}_{\mathrm{fs}_b,\mathrm{cs}_b}}$) and X_{29} (= $\mathbf{T}_{\mathrm{A}_{\mathbf{SRW}}\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}$), that as self-model states play the role of (adaptive) threshold combination function parameter for the excitability of the states X_5 (= bs_{Stigma}) and X_{22} (= $\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}$). So, for example, state X_5 (= bs_{Stigma}) has also a non-constant non-basic causal impact from that combination function parameter role, specified in role matrix **mcfp**.

Next, the *aggregation characteristics* of the adaptive network model are addressed. Four combination functions are used within the network: **alogistic**, **hebb**, **compid**, **stepmod**. In the combination function library used they have numbers 2, 3, 22, and 35, respectively, which is specified by

$$\mathbf{mcf} = \begin{bmatrix} 2 & 3 & 22 & 35 \end{bmatrix}$$

By specifying this **mcf**, within this specific network model these combination functions become number 1 to 4, respectively. Each of the 33 states gets one of

these combination functions assigned, which is specified in role matrix **mcfw** for combination function weights in Box 8.2. Most of them get **alogistic** assigned, which is often used as a kind of standard combination function. However, the three states X_{12}, X_{13}, X_{14} ($\mathbf{RW}_{\mathrm{srs}_{Complaints},\mathrm{bs}_{Stigma}}$, $\mathbf{RW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}$, $\mathbf{RW}_{\mathrm{srs}_{Complaints},\mathrm{bs}_{Irrelevant}}$ use the function **hebb**, which as a form of learning from experience allows them to acquire their values from simultaneous activation of the two connected states.

Note that, as these **RW**-states describe the person's knowledge or beliefs on the connection strengths; they are not used for the mental processing as those weights actually have the constant values 1 specified in role matrix **mcw** in Box 8.1, whereas the person's beliefs can concern different values. For example, if the person as a kind of blind spot about some effect, she or he can believe that the connection has a very low value (low value of the **RW**-state), whereas the value as specified in **mcw** is actually high. The two states X_{29}, X_{32} ($\mathbf{T}_{\mathbf{A}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}}$, $\mathbf{V}^{-}_{\mathbf{W}_{\mathrm{x},\mathrm{bs}_{Stigma}}}$) use the combination function **compid** which gives a complemental effect by mapping any incoming impact *V* to 1-*V*. Finally, the interventions by the therapist X_{18}, X_{27} ($\mathrm{int1}_{Therapist}$, $\mathrm{int2}_{Therapist}$) are considered external events for the person and therefore they use the combination function **stepmod** by which the timing of the event is specified.

Combination functions often have parameters, which form yet another type of the network's aggregation characteristics which exert causal impacts from a different role. These are specified in role matrix **mcfp** for combination function parameter values:

- Combination function **alogistic** has a *steepness* parameter **σ** and a *threshold* parameter **τ**.
- Combination function **hebb** has one parameter **μ** for the *persistence* rate.
- Combination function **stepmod** has a parameter **ρ** for *repetition* time of the event and **δ** for the *duration* of each occurrence of the event.

Note that the two adaptive threshold values for states X_5 (= bs_{Stigma}) and X_{22} (= $\mathbf{A}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}$) do not have a fixed value specified in role matrix **mcfp**, but instead their cells display a reference to another state, X_{33} (= $\mathbf{T}_{\mathrm{bs}_{Stigma}}$) resp. X_{29} (= $\mathbf{T}_{\mathbf{A}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}}$), that represents the dynamic value of this parameter. In this way, the adaptiveness of these network characteristics is specified explicitly within the model by using their self-model states; in the graphical connectivity pictures this relates to the downward pink arrows. For example, as a result of focusing attention on $\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}$ via

$\mathbf{F}_{\mathbf{SA}_{\mathbf{SRW}_{x,bs_{Stigma}}}}$, by X_{29} (= $\mathbf{T}_{\mathbf{A}_{\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}}}$) it is arranged that the threshold of $\mathbf{SA}_{\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}}$ is decreased, so that self-awareness shifts to it. After getting self-awareness $\mathbf{SA}_{\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}}$, the threshold $\mathbf{T}_{\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}}$ (=X_{25}) of self-referencing state $\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}$ is decreased, due to which by the awareness the latter state becomes more active and therefore also more accessible for any other state. Note that another, equally feasible, option to model this enhanced accessibility could be by increasing the weights of the outgoing connections from $\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}$.

	mcfw combination function weights	1 alog-istic	2 hebb	3 comp-id	4 step-mod
X_1	$int0_{Partner}$				1
X_2	$srs_{Complaints}$				1
X_3	ms_{Stigma}	1			
X_4	$ms_{Irrelevant}$	1			
X_5	bs_{Stigma}	1			
X_6	$bs_{Irrelevant}$	1			
X_7	$ds_{AvoidConfirmation}$	1			
X_8	ps_{GotoGP}	1			
X_9	$ps_{AvoidMentalHelp}$	1			
X_{10}	$\mathbf{W}ms_{Stigma},bs_{Stigma}$		1		
X_{11}	$\mathbf{W}ms_{Irrelevant},bs_{Irrelevant}$		1		
X_{12}	$\mathbf{RW}srs_{Complaints},bs_{Stigma}$		1		
X_{13}	$\mathbf{RW}ms_{Stigma},bs_{Stigma}$		1		
X_{14}	$\mathbf{RW}ms_{Irrelevant},bs_{Irrelevant}$		1		
X_{15}	$\mathbf{SRW}srs_{Complaints},bs_{Stigma}$	1			
X_{16}	$\mathbf{SRW}ms_{Stigma},bs_{Stigma}$	1			
X_{17}	$\mathbf{SRW}ms_{Irrelevant},bs_{Irrelevant}$	1			
X_{18}	$int1_{Therapist}$				1
X_{19}	$\mathbf{F}_{\mathbf{SRW}}x,bs_{Stigma}$	1			
X_{20}	$\mathbf{A}Comfortingideas$	1			
X_{21}	$\mathbf{SA}_{\mathbf{SRW}}srs_{Complaints},bs_{Stigma}$	1			
X_{22}	$\mathbf{SA}_{\mathbf{SRW}}ms_{Stigma},bs_{Stigma}$	1			
X_{23}	$\mathbf{SA}_{\mathbf{SRW}}ms_{Irrelevant},bs_{Irrelevant}$	1			
X_{24}	$\mathbf{T}_{\mathbf{SRW}}srs_{Complaints},bs_{Stigma}$			1	
X_{25}	$\mathbf{T}_{\mathbf{SRW}}ms_{Stigma},bs_{Stigma}$			1	
X_{26}	$\mathbf{T}_{\mathbf{SRW}}ms_{Irrelevant},bs_{Irrelevant}$			1	
X_{27}	$int2_{Therapist}$				1
X_{28}	$\mathbf{F}_{\mathbf{SA}_{\mathbf{SRW}}}x,bs_{Stigma}$	1			
X_{29}	$\mathbf{T}_{\mathbf{SA}_{\mathbf{SRW}}}ms_{Stigma},bs_{Stigma}$			1	
X_{30}	$\mathbf{V}^{-}{}_{\mathbf{SA}_{\mathbf{SRW}}}ms_{Stigma},bs_{Stigma}$	1			
X_{31}	$\mathbf{V}^{+}\mathbf{A}Comfortingideas$	1			
X_{32}	$\mathbf{V}^{-}\mathbf{W}x,bs_{Stigma}$	1			
X_{33}	$\mathbf{T}bs_{Stigma}$	1			

	mcfp combination function parameters	alogistic 1 σ	alogistic 2 τ	hebb 1 μ	hebb 2	compid 1	compid 2	stepmod 1 ρ	stepmod 2 δ
X_1	$int0_{Partner}$							120	40
X_2	$srs_{Complaints}$							120	20
X_3	ms_{Stigma}	15	0.1						
X_4	$ms_{Irrelevant}$	15	0.1						
X_5	bs_{Stigma}	5	X_{33}						
X_6	$bs_{Irrelevant}$	5	0.7						
X_7	$ds_{AvoidConfirmation}$	5	0.7						
X_8	ps_{GotoGP}	5	0.7						
X_9	$ps_{AvoidMentalHelp}$	5	1						
X_{10}	$\mathbf{W}ms_{Stigma},bs_{Stigma}$			1					
X_{11}	$\mathbf{W}ms_{Irrelevant},bs_{Irrelevant}$			1					
X_{12}	$\mathbf{RW}srs_{Complaints},bs_{Stigma}$			1					
X_{13}	$\mathbf{RW}ms_{Stigma},bs_{Stigma}$			1					
X_{14}	$\mathbf{RW}srs_{Complaints},bs_{Irrelevant}$			1					
X_{15}	$\mathbf{SRW}srs_{Complaints},bs_{Stigma}$	5	X_{24}						
X_{16}	$\mathbf{SRW}ms_{Stigma},bs_{Stigma}$	5	X_{25}						
X_{17}	$\mathbf{SRW}srs_{Complaints},bs_{Irrelevant}$	5	X_{26}						
X_{18}	$int1_{Therapist}$							120	60
X_{19}	$\mathbf{F}_{\mathbf{SRW}}x,bs_{Stigma}$	5	1.4						
X_{20}	$\mathbf{A}Comfortingideas$	5	0.8						
X_{21}	$\mathbf{SA}_{\mathbf{SRW}}srs_{Complaints},bs_{Stigma}$	5	0.65						
X_{22}	$\mathbf{SA}_{\mathbf{SRW}}ms_{Stigma},bs_{Stigma}$	5	X_{29}						
X_{23}	$\mathbf{SA}_{\mathbf{SRW}}ms_{Irrelevant},bs_{Irrelevant}$	5	0.5						
X_{24}	$\mathbf{T}_{\mathbf{SRW}}srs_{Complaints},bs_{Stigma}$								
X_{25}	$\mathbf{T}_{\mathbf{SRW}}ms_{Stigma},bs_{Stigma}$								
X_{26}	$\mathbf{T}_{\mathbf{SRW}}ms_{Irrelevant},bs_{Irrelevant}$								
X_{27}	$int2_{Therapist}$							120	80
X_{28}	$\mathbf{F}_{\mathbf{SA}_{\mathbf{SRW}}}x,bs_{Stigma}$	5	0.5						
X_{29}	$\mathbf{T}_{\mathbf{SA}_{\mathbf{SRW}}}ms_{Stigma},bs_{Stigma}$								
X_{30}	$\mathbf{V}^{-}{}_{\mathbf{SA}_{\mathbf{SRW}}}ms_{Stigma},bs_{Stigma}$	5	0.7						
X_{31}	$\mathbf{V}^{+}\mathbf{A}Comfortingideas$	5	0.7						
X_{32}	$\mathbf{V}^{-}\mathbf{W}x,bs_{Stigma}$	5	0.7						
X_{33}	$\mathbf{T}bs_{Stigma}$	10	0.5						

Box 8.2 Role matrices **mcfw** and **mcfp** for the aggregation characteristics of the adaptive network model

For the *timing characteristics* of the network model, another role matrix is used: the speed factor role matrix **ms** (see Box 8.3). These speed factors indicate how fast a state value changes upon causal impact received.

ms	**speed factors**	1
X_1	$int0_{Partner}$	2
X_2	$srs_{Complaints}$	1
X_3	ms_{Stigma}	1
X_4	$ms_{Irrelevant}$	1
X_5	bs_{Stigma}	0.25
X_6	$bs_{Irrelevant}$	0.25
X_7	$ds_{AvoidConfirmation}$	0.5
X_8	ps_{GotoGP}	0.5
X_9	$ps_{AvoidMentalHelp}$	0.5
X_{10}	$\mathbf{W}ms_{Stigma},bs_{Stigma}$	0.8
X_{11}	$\mathbf{W}ms_{Irrelevant},bs_{Irrelevant}$	0.8
X_{12}	$\mathbf{RW}srs_{Complaints},bs_{Stigma}$	0.3
X_{13}	$\mathbf{RW}ms_{Stigma},bs_{Stigma}$	0.3
X_{14}	$\mathbf{RW}srs_{Complaints},bs_{Irrelevant}$	0.3
X_{15}	$\mathbf{SRW}srs_{Complaints},bs_{Stigma}$	0.5
X_{16}	$\mathbf{SRW}ms_{Stigma},bs_{Stigma}$	0.5
X_{17}	$\mathbf{SRW}srs_{Complaints},bs_{Irrelevant}$	0.5
X_{18}	$int1_{Therapist}$	2
X_{19}	$\mathbf{F}_{SRW}x,bs_{Stigma}$	0.5
X_{20}	$\mathbf{A}_{Comfortingideas}$	0.5
X_{21}	$\mathbf{SA}_{SRW}srs_{Complaints},bs_{Stigma}$	0.2
X_{22}	$\mathbf{SA}_{SRW}ms_{Stigma},bs_{Stigma}$	0.2
X_{23}	$\mathbf{SA}_{SRW}ms_{Irrelevant},bs_{Irrelevant}$	0.2
X_{24}	$\mathbf{T}_{SRW}srs_{Complaints},bs_{Stigma}$	0.1
X_{25}	$\mathbf{T}_{SRW}ms_{Stigma},bs_{Stigma}$	0.1
X_{26}	$\mathbf{T}_{SRW}ms_{Irrelevant},bs_{Irrelevant}$	0.1
X_{27}	$int2_{Therapist}$	2
X_{28}	$\mathbf{F}_{SASRW}x,bs_{Stigma}$	0.8
X_{29}	$\mathbf{T}_{SASRW}ms_{Stigma},bs_{Stigma}$	0.8
X_{30}	$\mathbf{V}^{-}_{SASRW}ms_{Stigma},bs_{Stigma}$	0.8
X_{31}	$\mathbf{V}^{+}\mathbf{A}_{Comfortingideas}$	0.8
X_{32}	$\mathbf{V}^{-}\mathbf{W}x,bs_{Stigma}$	0.8
X_{33}	$\mathbf{T}bs_{Stigma}$	0.8

iv	**initial values**	1
X_1	$int0_{Partner}$	0
X_2	$srs_{Complaints}$	0
X_3	ms_{Stigma}	1
X_4	$ms_{Irrelevant}$	1
X_5	bs_{Stigma}	0
X_6	$bs_{Irrelevant}$	0
X_7	$ds_{AvoidConfirmation}$	0
X_8	ps_{GotoGP}	0
X_9	$ps_{AvoidMentalHelp}$	0
X_{10}	$\mathbf{W}ms_{Stigma},bs_{Stigma}$	0.1
X_{11}	$\mathbf{W}ms_{Irrelevant},bs_{Irrelevant}$	0.1
X_{12}	$\mathbf{RW}srs_{Complaints},bs_{Stigma}$	0.2
X_{13}	$\mathbf{RW}ms_{Stigma},bs_{Stigma}$	0.3
X_{14}	$\mathbf{RW}srs_{Complaints},bs_{Irrelevant}$	0.25
X_{15}	$\mathbf{SRW}srs_{Complaints},bs_{Stigma}$	0
X_{16}	$\mathbf{SRW}ms_{Stigma},bs_{Stigma}$	0
X_{17}	$\mathbf{SRW}srs_{Complaints},bs_{Irrelevant}$	0
X_{18}	$int1_{Therapist}$	0
X_{19}	$\mathbf{F}_{SRW}x,bs_{Stigma}$	0
X_{20}	$\mathbf{A}_{Comfortingideas}$	0
X_{21}	$\mathbf{SA}_{SRW}srs_{Complaints},bs_{Stigma}$	0
X_{22}	$\mathbf{SA}_{SRW}ms_{Stigma},bs_{Stigma}$	0
X_{23}	$\mathbf{SA}_{SRW}ms_{Irrelevant},bs_{Irrelevant}$	0
X_{24}	$\mathbf{T}_{SRW}srs_{Complaints},bs_{Stigma}$	1
X_{25}	$\mathbf{T}_{SRW}ms_{Stigma},bs_{Stigma}$	1
X_{26}	$\mathbf{T}_{SRW}ms_{Irrelevant},bs_{Irrelevant}$	0.5
X_{27}	$int2_{Therapist}$	0
X_{28}	$\mathbf{F}_{SASRW}x,bs_{Stigma}$	0
X_{29}	$\mathbf{T}_{SASRW}ms_{Stigma},bs_{Stigma}$	1
X_{30}	$\mathbf{V}^{-}_{SASRW}ms_{Stigma},bs_{Stigma}$	0
X_{31}	$\mathbf{V}^{+}\mathbf{A}_{Comfortingideas}$	0
X_{32}	$\mathbf{V}^{-}\mathbf{W}x,bs_{Stigma}$	0
X_{33}	$\mathbf{T}bs_{Stigma}$	0

Box 8.3 Role matrix **ms** for the timing characteristics of the adaptive network model and the initial values **iv**

8.7 Example Simulation for the Case Study

Based on the specification of the network model shown in Boxs 8.1, 8.2 and 8.3 in Sect. 8.6 and the dedicated software environment described in (Treur 2020b, Ch 9), a simulation has been generated for the case study. The overall outcome is shown in Fig. 8.8. The vertical lines indicate that intervention 0 by the partner took place at time 40, intervention 1 by the therapist at time 60 and intervention 2 at time 80. Moreover, at time 20 the complaints start. It can be seen that intervention 0 only has a minor effect and intervention 2 has a major effect. However, with 33 lines in one graph, this is not easy to understand in more detail. Therefore, in the next 4 Figs. 8.9 , 8.10, 8.11 and 8.12, parts are shown.

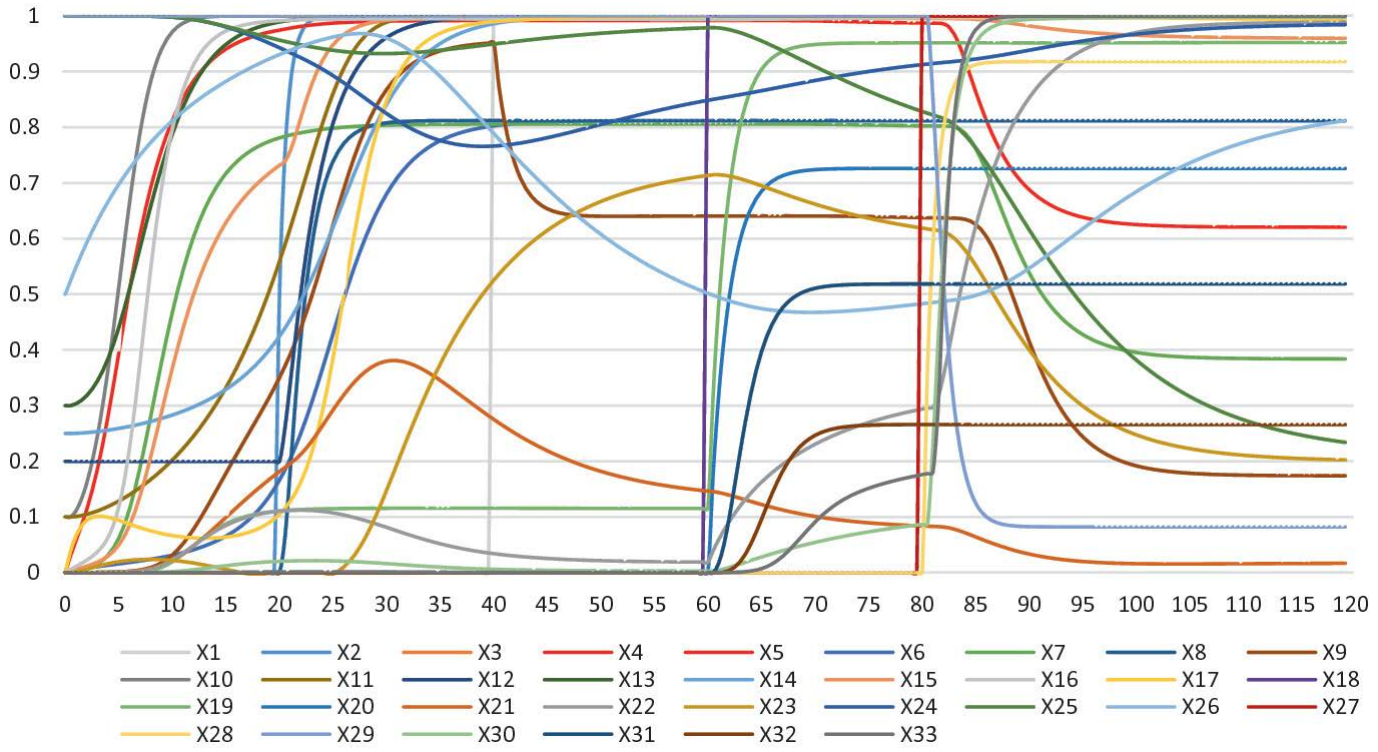


Fig. 8.8 Overall outcome of the simulation

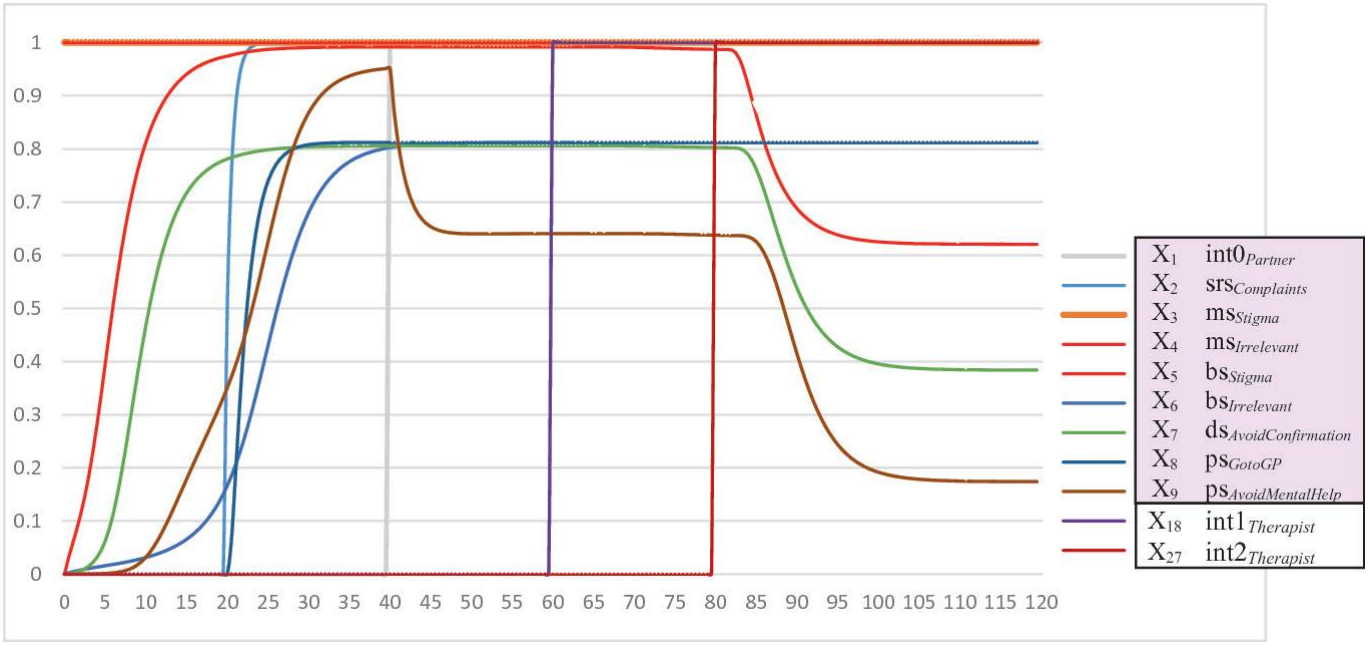


Fig. 8.9 The interventions and the effects for the base states

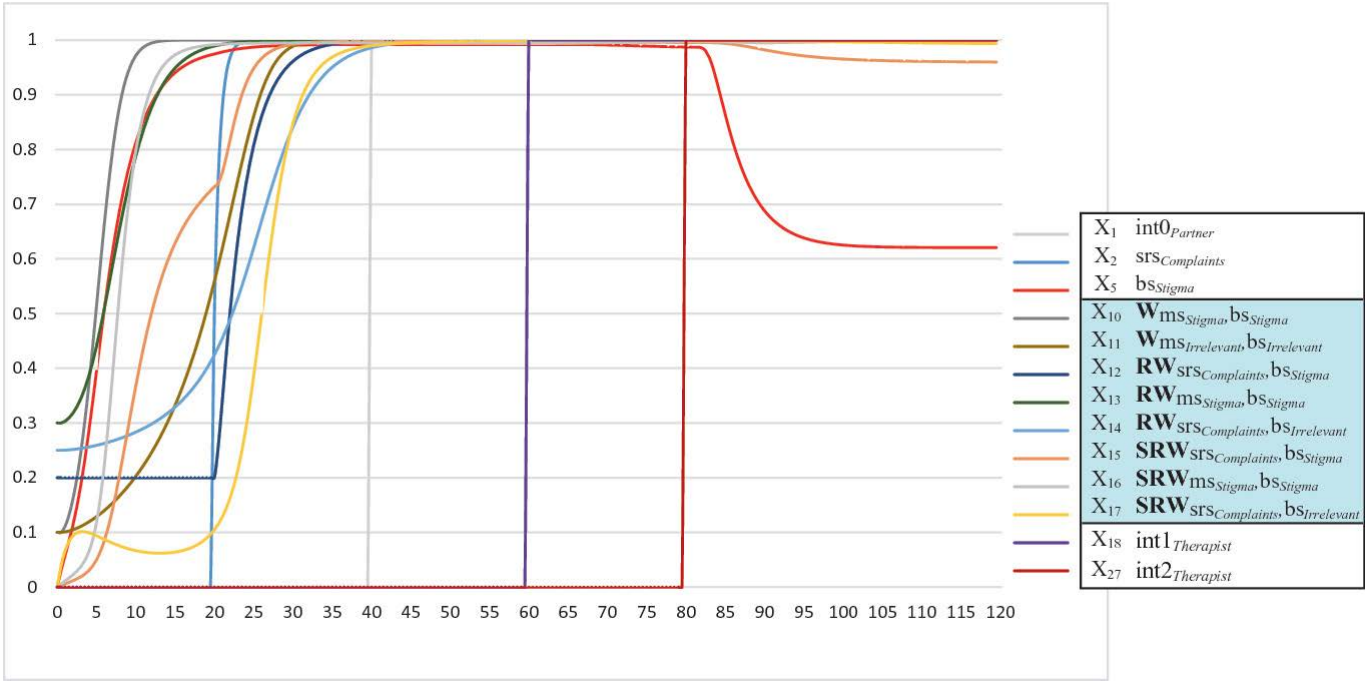


Fig. 8.10 The effects for the self-referencing states

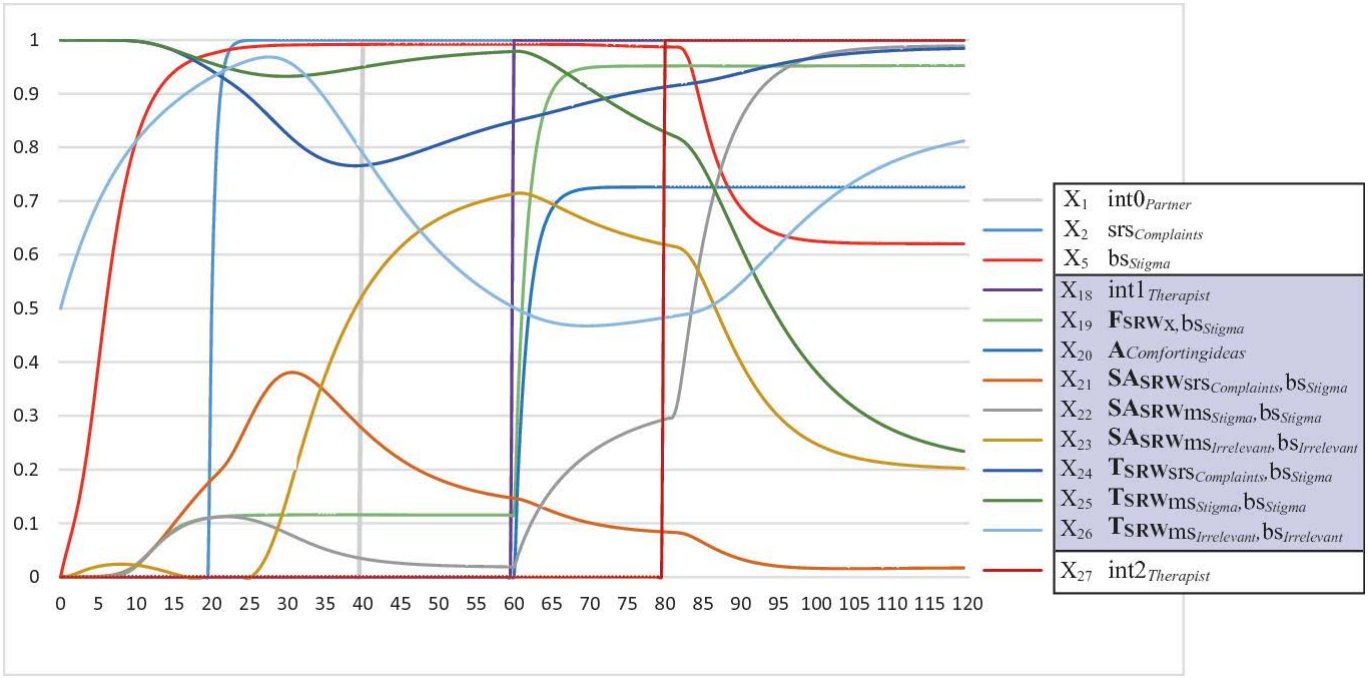


Fig. 8.11 The effects for the self-awareness states

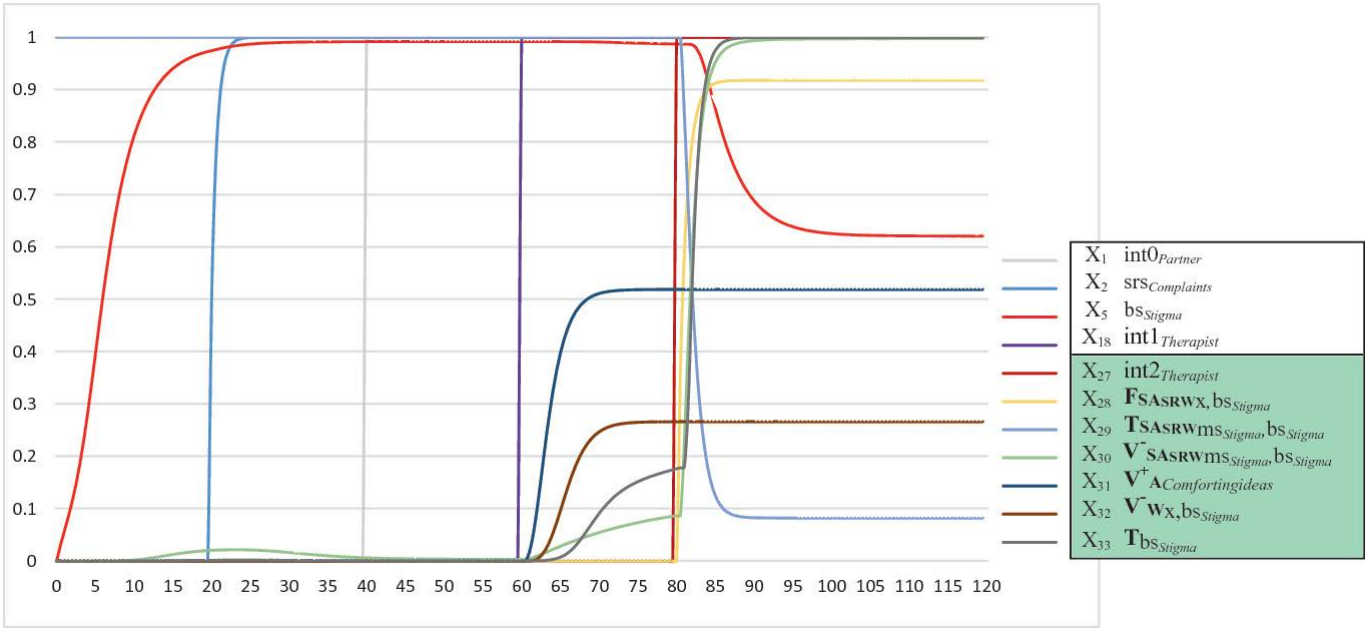


Fig. 8.12 The effects for the self-interpretation states

In Fig. 8.9 the focus is on the base states, with also the interventions visible. After the complaints start at time 20, the preparation to go to a GP becomes high (the dark blue line). It can also be seen that after intervention 0 by the partner the being prepared not to go for mental help (the brown line) decreases until below 0.7, after intervention 1 the level of the belief state bs_{Stigma} decreases very lightly (the red line), but still stays very high. However, after intervention 2 the belief state bs_{Stigma} goes down from above 0.95 to just above 0.6. Hand in hand with this also the desire state $ds_{AvoidConfirmation}$ (the green line) and the

preparation state $ps_{AvoidMentalHelp}$ (the brown line) go down to below 0.4 and 0.2, respectively.

To see how these effects were achieved, in the next figures we zoom in on the higher level states, first in Fig. 8.10 on the self-referencing states covering of the first-order internal self-model. In the first phase between 0 and 20 it is shown how the connections to the belief states bs_{Stigma} and $bs_{Irrelevant}$ are learned, shown by their self-model states $\mathbf{W}_{ms_{Stigma},bs_{Stigma}}$ and $\mathbf{W}_{ms_{Irrelevant},bs_{Irrelevant}}$ (dark green line and brown line both starting at 0.1). It can be seen that initially already representations $\mathbf{RW}_{srs_{Complaints},bs_{Stigma}}$, $\mathbf{RW}_{ms_{Complaints},bs_{Irrelevant}}$ and $\mathbf{RW}_{ms_{Stigma},bs_{Stigma}}$, were present of the connections, with activation levels 0.2, 0.25, and 0.3, but these do not correspond exactly to the actual values which are higher: the person has a first-order model on the connections, but, as is the case more often such a model is not exactly the reality. However, due to experiencing their effects, these values finally increase to 1, what is indeed their actual value. As belief state bs_{Stigma} also becomes high (red line), together with the above **RW**-states this activates the self-referencing states $\mathbf{SRW}_{srs_{Complaints},bs_{Stigma}}$, $\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}$, $\mathbf{SRW}_{srs_{Complaints},bs_{Irrelevant}}$ (initially all 0) which relates the occurrence of belief states bs_{Stigma} and $bs_{Irrelevant}$ to their causing state ms_{Stigma}.

All of the above states do not involve self-awareness yet. The self-awareness states are described in Fig. 8.11. In the first phase there is practically no self-awareness $\mathbf{SA}_{\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}}$ of the role of the memory state ms_{Stigma} in causing belief state bs_{Stigma} (see the grey line). During this phase, self-awareness focuses more on the belief $bs_{Irrelevant}$ (the light brown line peaking just above 0.7 around time 60). Being awareness-states, these self-awareness states mutually suppress each other, where apparently $\mathbf{SA}_{\mathbf{SRW}_{ms_{Complaints},bs_{Irrelevant}}}$ is the first winner of the competition. Intervention 1 at time 60 brings a focus state $\mathbf{F}_{\mathbf{SRW}_{x,bs_{Stigma}}}$ (light green line) that results in a small but still insufficient change for the awareness up till time 80.

However, intervention 2 at time 80 brings about a more serious change, so that after that $\mathbf{SA}_{\mathbf{SRW}_{ms_{Stigma},bs_{Stigma}}}$ becomes the winner of the competition (the steep upward grey line). It can be seen that the strong increase of the latter state goes together with a substantial decrease for the belief state bs_{Stigma}. How that can happen is explained by the next graph in Fig. 8.10 involving self-interpretation.

In Fig. 8.12 it can be seen that at this self-interpretation level, by intervention 2 many things change strongly. First, the focus $\mathbf{F}_{\mathbf{SA}_{\mathbf{SRW}_{x,bs_{Stigma}}}}$ on the causes of

belief state bs_{Stigma} initiated by the therapist becomes active (the yellow line, ending above 0.9). This leads to a strong decrease of state $\mathbf{T}_{\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}}$ (the steep downward blue-grey line) indicating the (previously high) excitability threshold which was blocking until then the rise of the awareness state for $\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}$. As this threshold becomes low, $\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}$ now gets a chance to increase, which indeed happens as already was observed in Fig. 8.11, grey line. Moreover, as part of the analysis for self-interpretation, a negative valuation $\mathbf{V}^{-}_{\mathbf{SA}_{\mathbf{SRW}_{\mathrm{ms}_{Stigma},\mathrm{bs}_{Stigma}}}}$ develops (the steep upward light green line). Through this negative valuation, the person generates the state $\mathbf{T}\mathrm{bs}_{Stigma}$ (steep upward dark green line, ending at 1) meant to help block or at least decrease the belief state bs_{Stigma} as a form of self-control. This is indeed the reason that the belief state goes down in this stage (through the long pink downward link in Fig. 8.7), which explains what was already observed in the previous figures.

8.8 Discussion

In this chapter, a multilevel cognitive architecture is introduced that can be used to model mental processes in clients of psychotherapeutic sessions. The architecture does not only cover base level mental processes but also mental processes involving self-referencing, self-awareness and self-interpretation; e.g., (Glas 2017, 2019; Glass 2020). To this end, the cognitive architecture was designed according to four levels, where (part of) the structure of each level is represented by an explicit mental self-model of it at the next-higher level of the architecture. At that next-higher level, states represent this structure and have a referencing relation to it. Most material is based on (Treur and Glas 2021). The self-modeling network model shows how a person can have mental self-models of him- or herself of different levels of description.

The cognitive architecture was evaluated for a case study of a realistic type of a therapeutic session from clinical practice. The relevant aspects of this case study concerning self-referentiality, self-awareness and self-interpretation were all covered well by applying the architecture.

The cognitive architecture was specified formally using the role matrix language format for self-modeling networks described in (Treur 2020b), which extends the approach introduced in (Treur 2016a, b) by adding self-models and a systematic approach to (multi-order) adaptive networks, introduced in (Treur 2018; Treur 2020a). Using the available dedicated software environment, the simulations were conducted automatically with this formal specification as input.

It has been shown how by the cognitive architecture self-awareness states can be modeled, taking into account three well-known principles occurring in multiple theories of consciousness: the role of attention, a winner-takes-it-all principle, and enhanced accessibility; e.g., (Graziano 2013; Graziano et al. 2019; Baars 1997). In addition, the leveled structure of the cognitive architecture has some relation to higher-order theories of consciousness; e.g., (Metzinger 2003, 2007; Rosenthal 2005).

In future work, a number of issues can be addressed further. One of them is to model examples of therapeutic sessions in which emotions play a main role; e.g., (Glas 2020). Also other case studies from clinical practice will be considered for further evaluation. Further future development may consider to use the architecture for virtual training situations for therapists.

References

Atkinson, A.O., Ratcliffe, M.: Introduction to the special section on "emotions and feelings in psychiatric illness." Emot. Rev. **4**(2), 119–121 (2012)
[Crossref]

Baars, B.J.: In the Theater of Consciousness: The Workspace of the Mind. Oxford University Press, Oxford (1997)
[Crossref]

Bowen, K.A., Kowalski, R.: Amalgamating language and meta-language in logic programming. In: Clark, K., Tarnlund, S. (eds.) Logic Programming, pp. 153–172. Academic Press, New York (1982)

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

Christoff, K., Cosmelli, D., Legrand, D., Thompson, E.: Specifying the self for cognitive neuroscience. Trends Cogn. Sci. **15**(3), 104–112 (2011)
[Crossref]

Damasio, A.R.: Self Comes to Mind: Constructing the Conscious Brain. Pantheon Books, New York (2010)

Damasio, A.R.: The Feeling of What Happens. Body and Emotion in the Making of Consciousness. San Diego: Harcourt (1999)

Descartes, R.: Principles of Philosophy (translated by MS Mahoney). (1644)

Fessler, D.M.T., Eng, S.J., Navarrete, C.D.: Elevated disgust sensitivity in the first trimester of pregnancy: evidence supporting the compensatory prophylaxis hypothesis. Evol. Hum. Behav. **26**(4), 344–351 (2005)
[Crossref]

Fessler, D.M.T., Clark, J.A., Clint, E.K.: Evolutionary psychology and evolutionary anthropology. In: Buss, D.M. (ed.) The Handbook of Evolutionary Psychology, pp. 1029–1046. Wiley (2015)

Fonagy, P., Gergely, G., Jurist, E.L., Target, M. (eds.): Affect Regulation, Mentalization, and the Development of the Self. Other Press, New York (2002)

Gallagher, S.: A pattern theory of self. Front. Hum. Neurosci. **7**, e443, 1–7 [online: 01 August 2013 doi: https://doi.org/10.3389/fnhum.2013.00443] (2013)

Galton, A.: Operators versus arguments: the ins and outs of reification. Synthese 150, 415–441

Gascoigne, N., Thornton, T.: Tacit Knowledge. Routledge, London & NY (2014). (2006)

Glas, G.: Dimensions of the self in emotion and psychopathology. Philos. Psychiatry Psychol. **24**(2), 143–155 (2017)
[MathSciNet][Crossref]

Glas, G.: An enactive approach to anxiety and anxiety disorders. Philos. Psychiatry Psychol. **27**(1), 35–50 (2020)
[Crossref]

Glas, G.: Person-Centered Care in Psychiatry. Selfrelational, Contextual, and Normative Perspectives. Routledge, Abingdon/London (2019)

Graziano, M.S.A.: Consciousness and the Social Brain. Oxford University Press, New York (2013)

Graziano, M.S.A., Guterstam, A., Bio, B.J., Wilterson, A.I.: Toward a standard model of consciousness: reconciling the attention schema, globalworkspace, higher-order thought, and illusionist theories. Cogn. Neuropsychol. **37**(3–4), 155–172 (2019). https://doi.org/10.1080/02643294.2019.1670630
[Crossref]

Gross, J.J., Thompson, R.A.: Emotion regulation: conceptual foundations. In: Gross, J.J. (ed.) Handbook of Emotion Regulation, pp. 3–24. Guilford, New York (2007)

Gross, J.J., Jazaieri, H.: Emotion, emotion regulation, and psychopathology: an affective science perspective. Clin. Psychol. Sci. **2**, 387–401 (2014)
[Crossref]

Gross, J.J.: Emotion regulation. In: Lewis, M., Haviland-Jones, J.M., Barrett, L.F. (eds.) Handbook of Emotions, 3rd edn, pp. 497–512. Guilford, New York (2008)

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

Hobson, P.: Emotion, self/other-awareness, and autism: a developmental perspective. In: Goldie, P. (ed.) The Oxford Handbook of Philosophy and Emotion, pp. 445–472. Oxford: Oxford University Press (2010)

Iacoboni, M.: Mirroring People: The New Science of How We Connect with Others. Farrar, New York (2008)

Immordino-Yang, M.H.: Me, My "Self" and You: neuropsychological relations between social emotion, self-awareness, and morality. Emot. Rev. **3**(3), 313–315 (2011)
[Crossref]

Kappas, A.: Emotion and regulation are one! Emot. Rev. **3**(1), 17–25 (2011)
[Crossref]

Keysers, C., Gazzola, V.: Hebbian learning and predictive mirror neurons for actions, sensations and emotions. Philos. Trans. r. Soc. Lond. B Biol. Sci. **369**, 20130175 (2014)
[Crossref]

Kim, J.: Philosophy of Mind. Westview Press (1996)

Koole, S.: The psychology of emotion regulation: an integrative review. Cogn. Emot. **23**(1), 4–41 (2009)
[Crossref]

Leary, M.R.: Motivational and emotional aspects of the self. Annu. Rev. Psychol. **58**, 317–344 (2007)
[Crossref]

Leary, M.R., Tangney, J.P.: The self as an organizing construct in the behavioural and social sciences. In: Leary M.R., Tangney J.P. (eds) Handbook of Self and Identity, pp. 3–14. The Guilford Press, New York/London (2003)

Leibniz, G.W.: Phoranomus seu De potentia et legibus naturæ. Dialogus II Republished in (1991) in: Physis **28**, 797–885 (1698)

Levy, D.A., Nail, P.R.: Contagion: a theoretical and empirical review and reconceptualization. Genet. Soc. Gen. Psychol. Monogr. **119**(2), 233–284 (1993)

Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**, 130–141 (1963)
[MathSciNet][Crossref]

Lorenz, E.N.: The Essence of Chaos. University of Washington Press (1993)

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

Metzinger, T.: Being No One: The Self-model Theory of Subjectivity. MIT Press (2003)

Metzinger, T.: Empirical perspectives from the self-model theory of subjectivity: a brief summary with examples. In: Banerjee, R., Chakrabarti, B.K. (eds.) Models of Brain and Mind: Physical, Computational and Psychological Approaches. Progress in Brain Research, vol. 168, pp. 215–245, 273–278. Elsevier (2007)

Minsky, M.: The Society of Mind. Simon & Schuster, New York (1986)

Montgomery, K.: How Doctors Think: Clinical Judgment and the Practice of Medicine. Oxford University Press, Oxford (2006)

Nagel, E., Newman, J.: Gödel's Proof. New York University Press, New York (1965)
[zbMATH]

Newton, I.: The Mathematical Principles of Natural Philosophy; Newton's Principles of Natural Philosophy, Dawsons of Pall Mall, 1968 (1729)

Nicolini, D.: Practice Theory, Work, and Organization. An Introduction. Oxford University Press, Oxford (2012)

Northoff, G., Qin, P., Feinberg, T.E.: Brain imaging of the self-conceptual, anatomical and methodological issues. Conscious. Cogn. **20**, 52–63 (2011)
[Crossref]

Prinz, J.: The moral emotions. In: Goldie P. (ed.) The Oxford Handbook of Philosophy and Emotion, pp. 519–538. Oxford: Oxford University Press (2010)

Ratcliffe, M.: The feeling of being. J. Conscious. Stud. **12**, 43–60 (2005)

Ratcliffe, M.: Feelings of Being. Phenomenology, Psychiatry and the Sense of Reality. Oxford: Oxford University Press (2008)

Ratcliffe, M.: The phenomenology of mood and the meaning of life. In: P. Goldie (ed.) The Oxford Handbook of Philosophy and Emotion, pp. 349–372. Oxford: Oxford University Press (2010).

Reddy, W.M.: Historical research on the self and emotions. Emot. Rev. **1**(4), 302–315 (2009)
[Crossref]

Ricoeur, P.: Oneself as Another (trans. Kathleen Blamey). University of Chicago, Chicago (original work published in 1990) (1992)

Rorty, A.: A plea for ambivalence. In: Goldie, P. (ed.) The Oxford Handbook of Philosophy and Emotion, pp. 425–444. Oxford University Press, Oxford (2010)

Rosenthal, D.: Consciousness and Mind. Oxford University Press, Oxford (2005)

Slaby, J., Stephan, A.: Affective intentionality and self-consciousness. Conscious. Cogn. **17**, 506–513 (2008) [Crossref]

Smorynski, C.: The incompleteness theorems. In: Barwise, J. (ed.) Handbook of Mathematical Logic, vol. 4, pp. 821–865. North-Holland, Amsterdam (1977)

Solomon, R.C.: The Passions. University of Notre Dame Press, Notre Dame (1983)

Stephan, A.: Emotions, existential feelings, and their regulation. Emot. Rev. **4**(2), 157–162 (2012) [Crossref]

Sterling, L., Beer, R.: Metainterpreters for expert system construction. J. Log. Program. **6**, 163–178 (1989) [Crossref]

Stern, D.N.: The Present Moment in Psychotherapy and Everyday Life. Norton, New York & London (2004)

Tracy, J.L., Robins, R.W.: The self in social psychology. In: C. Sedikides & S. Spence, (eds.) Frontiers of Social Psychology Series, pp. 187–209. Psychology Press, New York (2007)

Tracy, J.L., Klonsky, E.D., Proudfit, G.H.: How affective science can inform clinical science. An introduction to the special series on emotions and psychopathology. Clin. Psychol. Sci. **2**(4), 371–386 (2014)

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

Treur, J.: Dynamic modeling based on a temporal–causal network modeling approach. Biol. Inspired Cogn. Archit. **16**, 131–168 (2016a)

Treur, J.: Network-Oriented Modeling: Addressing Complexity of Cognitive, Affective and Social Interactions. Springer Nature Publishers, Cham, Switzerland (2016b)

Treur, J.: Network reification as a unified approach to represent network adaptation principles within a network. In: Martín-Vide C, Vega-Rodríguez MA, Fagan D, O'Neill M (eds) Proceedingsof the 7th International Conference on Theory and Practice of Natural Computing, TPNC'18. Lecture Notes in Computer Science, vol 11324, pp. 344–358. Springer Nature Publishers (2018)

Treur, J.: Modeling higher-order adaptivity of a network by multilevel network reification. Netw. Sci. **8**(S1), S110–S144 (2020a)

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

Tse, P.U.: The Neural Basis of Free Will: Criterial Causation. MIT Press, Cambridge (2013) [Crossref]

Westerhoff, H.V., He, F., Murabito, E., Crémazy, F., Barberis, M.: Understanding principles of the dynamic biochemical networks of life through systems biology. In: Kriete, A., Eils, R. (eds.) Computational Systems Biology, 2nd edn, pp. 21–44. Academic Press, Oxford (2014a)

Westerhoff, H.V., Brooks, A.N., Simeonidis, E., García-Contreras, R., He, F., Boogerd, F.C., Jackson, V.J., Goncharuk, V., Kolodkin, A.: Macromolecular networks and intelligence in microorganisms. Front. Microbiol. **5**, e379 (2014b)

Weyhrauch, R.W.: Prolegomena to a theory of mechanized formal reasoning. Artif. Intell. **13**, 133–170 (1980) [MathSciNet][Crossref]

Zahavi, D.: Subjectivity and Selfhood. Investigating the First-Person Perspective. MIT Press, Cambridge (2005) [Crossref]

Part III
Self-Modelling Network Models for Mental Models in Social Processes