



Acquiring Contextualized Concepts: A Connectionist Approach

Saskia van Dantzig,^a Antonino Raffone,^{a,b} Bernhard Hommel^a

^a*Department of Psychology, Leiden University*

^b*Department of Psychology, Sapienza University of Rome*

Received 23 November 2009; received in revised form 8 November 2010; accepted 10 November 2010

Abstract

Conceptual knowledge is acquired through recurrent experiences, by extracting statistical regularities at different levels of granularity. At a fine level, patterns of feature co-occurrence are categorized into objects. At a coarser level, patterns of concept co-occurrence are categorized into contexts. We present and test CONCAT, a connectionist model that simultaneously learns to categorize objects and contexts. The model contains two hierarchically organized CALM modules (Murre, Phaf, & Wolters, 1992). The first module, the *Object Module*, forms object representations based on co-occurrences between features. These representations are used as input for the second module, the *Context Module*, which categorizes contexts based on object co-occurrences. Feedback connections from the Context Module to the Object Module send activation from the active context to those objects that frequently occur within this context. We demonstrate that context feedback contributes to the successful categorization of objects, especially when bottom-up feature information is degraded or ambiguous.

Keywords: Concept learning; Top-down context influence; Hierarchical categorization; Neural network

1. Acquiring contextualized concepts: A connectionist approach

Whenever we interact with a particular object, we automatically activate stored knowledge about similar objects that we have encountered before. In other words, we activate a *concept*, a mental representation referring to a particular class of objects in the world. Concepts enable us to recognize and categorize objects, to draw inferences about their

Correspondence should be sent to Saskia van Dantzig, Department of Psychology, Leiden University, Wassenaarseweg 52, Room 2B11, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: saskia@vandantzig.nl

characteristics, and to react to them appropriately. Moreover, we can think and talk about objects that are not directly present by using our conceptual knowledge of those objects. Thus, concepts are the basic building blocks of cognition, and they are necessary to understand and interact with the world around us (Murphy, 2004).

Most theories of concept learning assume that conceptual knowledge of objects is acquired through experience (e.g., O'Connor, Cree, & McRae, 2009; Rogers & McClelland, 2004; Smith, Shoben, & Rips, 1974). During recurrent interactions with different exemplars of a class of objects (e.g., balls), the features that are invariant across those exemplars (e.g., shape), and thus relevant for the category, are extracted and stored in memory to form a concept of the category (the concept BALL¹). The features that vary across exemplars (e.g., color) are discarded, as well as information about the different contexts in which exemplars are encountered. These models thus assume that concepts are primarily based on bottom-up feature information, and they generally ignore the role of context information in conceptual processing (Yeh & Barsalou, 2006).

However, across many domains of cognition there is evidence that the context profoundly influences conceptual processing and performance (for an overview, see Yeh & Barsalou, 2006). Studies of visual processing have shown that visual recognition of objects is influenced by context information (for an overview see Bar, 2004). For example, Davenport and Potter (2004) found that pictures of objects (e.g., a camel) are recognized more accurately when they are seen in front of a typical background setting (e.g., a desert) than when they are seen in isolation or in front of an atypical background (e.g., a library). Objects are also recognized better when they are presented together with a contextually related object than with an unrelated object (Davenport, 2007). These effects are not limited to visual objects. A recent study by Özcan and van Egmond (2009) showed that sounds of household products (e.g., an electric toothbrush) are recognized faster and more accurately when congruent visual context information (e.g., a picture of a bathroom) is presented than when incongruent information (e.g., a picture of a kitchen) is presented. Context effects were also found in language processing (e.g., Potter, 1993; Potter, Kroll, Yachzel, Carpenter, & Sherman, 1986; Schwanenflugel & Shoben, 1985; Zeelenberg, Pecher, Shiffrin, & Raaijmakers, 2003) and in memory tasks, affecting both true memories (e.g., Godden & Baddeley, 1975; Mandler & Parker, 1976; Mandler & Stein, 1974) and false memories (Miller & Gazzaniga, 1998).

Furthermore, context effects were found in many conceptual tasks. Barsalou (1982) demonstrated that mental concepts are flexible and that their exact content depends on situational factors such as current goals and task demands. In a property-verification task, participants were faster to verify contextually relevant than contextually irrelevant properties. Other studies showed that typicality ratings of category instances depend on the context in which the instances are presented (e.g., Barsalou, 1982; Hampton, Dubois, & Yeh, 2006). Finally, context information can contribute to the categorization of novel objects. A recent study by Chaigneau, Barsalou, and Zamani (2009) showed that novel objects were categorized more accurately when context information was provided than when the objects were presented in isolation. In addition, inferences about the objects became more accurate as more context information was presented during categorization.

Context information thus appears to be very relevant for conceptual processing and for mechanisms that rely on conceptual representations. To account for this influence, we propose a connectionist model of concept learning, called CONCAT. This model is based on the assumption that the conceptual system, while learning object representations, does not ignore information about the different contexts in which these objects appear, but instead uses this information to form *context representations*. This aspect distinguishes CONCAT from existing models of concept learning, which typically assume that context information is ignored during the process of learning object representations. We assume that object representations and context representations are acquired in similar ways: by extracting the statistical regularities that emerge from repeated experiences. Such regularities can be detected at various temporal and spatial scales. At a fine scale, patterns of co-occurrence between low-level sensory-motor features (e.g., <red>, <round>, <soft>, <bouncing>) emerge, because the features belonging to a particular object tend to be grouped in time and space. These patterns of feature co-occurrence can be used to form object representations. At a coarser scale, patterns of co-occurrence between objects emerge. Objects are not distributed randomly in the world but tend to co-occur in time and space with other objects, resulting in coherent scenes or contexts. For example, fridges are typically found in kitchens and tend to co-occur with other objects that are often found in kitchens, such as ovens, stoves, sinks, and pans. A context can thus be defined as a *coherent collection of objects that co-occur within a certain window of time and space*. Thus, a room is defined as a kitchen by the objects that are typically found in kitchens. The idea that a context can be described as a pattern of co-occurring objects is not entirely new. For instance, Bar (2004) proposed the term “context frame” as a representation of a prototypical scene, containing information about the objects that are most likely to appear in this scene. Likewise, Tversky, Zacks, and Martin (2008) argued that scenes (contexts) are characterized by the activities that take place in them and the objects that occur in them. These objects are not distributed randomly across a scene, but their spatial arrangement is loosely constrained by the scene (see also Bar, 2004; Biederman, Mezzanotte, & Rabinowitz, 1982). For example, in a dining room, chairs are typically grouped around a table, a lamp is usually hanging above the table and food is placed on the table, rather than on the chairs. Although highly informative, spatial information is not used in our model. For the current purposes, a context is defined only as a pattern of co-occurrence between objects, disregarding information about their typical spatial arrangements.

In the model, associative connections exist between object representations and context representations. Activating a particular object representation will lead to activation of the context representations that are associated with this object. In turn, activating a context will prime objects that are likely to occur within this context. As a result of the connections between context representations and object representations, the latter contain not only feature information but also information about the contexts in which an object typically occurs. This is in line with the results of a property norming study of McRae, Cree, Seidenberg, and McNorgan (2005), who collected semantic properties for a large number of living (e.g., DOG) and nonliving objects (e.g., CHAIR). Participants were asked to list as many properties as possible for each object. Interestingly, participants not only provided sensory-motor

features such as <red> and <round>, but they also mentioned the contexts in which objects are typically found. For example, <found in kitchens> was often provided for the concept KNIFE. This shows that context information is an intrinsic and relevant part of object representations, which spontaneously comes to mind when people are thinking about a particular object.

Thus, object representations are based on two sources of information: feature-based or *experiential* data, and context-based or *distributional* data.² With respect to the former, we are basing our reasoning on the Theory of Event Coding (TEC; Hommel, Müsseler, Aschersleben, & Prinz, 2001), which assumes that stimulus and action events are cognitively represented in terms of their distal features, such as shape, size, distance, and location, rather than to proximal features of the sensations elicited by stimuli (e.g., retinal location) or of the motor patterns of an action (e.g., the contraction of specific muscles). For example, a left-hand response and a stimulus on the left both activate the same distal code “left.” We take feature-based information to refer to all the features (also called properties or attributes) of a particular category that are acquired through interactions with different exemplars of that category. This includes perceptual information from various sensory modalities, referring to the category’s typical shape, color, smell, taste, and feel but also affective information and information about the actions afforded by (exemplars of) the given category. The similarity between two concepts is partially defined by the degree to which their features overlap (see also McRae, 2004). In addition to this experiential information, object representations are based on distributional information, which refers to the different contexts in which an object is encountered (see also Barsalou & Wiemer-Hastings, 2005). Objects may occur in a range of different contexts. The similarity between two objects is therefore also defined by the degree to which they share the same contexts.

Although we focus only on object representations and context representations, we acknowledge that other levels of representation could be distinguished as well. For example, concepts can represent low-level perceptual features (e.g., RED), object parts (e.g., LEG), actions (e.g., THROW), and events (e.g., WEDDING). These concepts differ in the level of granularity, and each concept can encompass other concepts. Thus, representations are formed at multiple hierarchical levels, and each level essentially forms the context for the level below. For the sake of parsimony, we distinguish only the two levels mentioned above. Importantly, the suggested model can easily be applied to object configurations, actions, and other events, and it can be extended to capture hierarchical relations between these entities.

Summarizing, we assume that object representations and context representations are acquired over the course of multiple experiences, by extracting statistical regularities at various temporal and spatial scales. At a small scale, features that co-occur in time and space are bound into object representations. At a coarser scale, objects that co-occur in time and space are bound into context representations. Object representations and context representations are connected to each other, such that a particular object will activate the contexts in which it frequently appears and reversely, a particular context will activate the objects that are likely to occur within this context. Object representations are based on two sources of information; *experiential data*, bottom-up information about the (sensory, motor, affective, etc.) features that are relevant for interaction with the object, and *distributional data*,

top-down information about the different contexts in which the object is likely to occur. These functional principles are implemented in a connectionist model, named CONCAT, which simultaneously learns to categorize objects and contexts.

2. The model

2.1. Model overview, architecture, and functional logic

At the core of the CONCAT model are two Categorization and Learning Modules (CALM; Murre et al., 1992; Murre, 1992). A CALM module is able to perform unsupervised categorization of distributed input patterns in a localist fashion (see also Page, 2000). CALM modules display two modes of learning, depending on the novelty of the input pattern (Graf & Mandler, 1984; Wolters & Phaf, 1990). When a novel input pattern is presented, *elaboration learning* takes place. An increased learning rate combined with the distribution of nonspecific, random activations in the module results in strong competition between the uncommitted nodes (those nodes that do not yet represent a pattern). As a result, the input pattern is associated with a node that is not yet committed by any other pattern. When a familiar input pattern is presented, *activation learning* takes place. A low learning rate and weak competition between nodes result in the activation and strengthening of the existing representation. Multiple CALM modules can be combined to construct a modular neural network that is able to perform hierarchical categorization. CALM networks thus contain some psychologically and biologically plausible properties, such as modularity, unsupervised learning and novelty-dependent learning (Murre et al., 1992).

In CONCAT, depicted in Fig. 1, two CALM modules are implemented to perform a hierarchical categorization of input patterns. The first CALM module, the *Object Module*, uses feature co-occurrences to form object representations. These representations are used as input for the second CALM module, the *Context Module*, which forms context representations based on object co-occurrences. The CONCAT model incorporates two activation streams; a bottom-up or feedforward stream and a top-down or feedback stream. The feedforward stream originates at the input layer, called *Feature Detection Layer*. In this layer, a perceived object is represented as a distributed pattern of activation across nodes representing different features from various sensory modalities, as well as affective and affordance features. We assume that some preprocessing of the raw sensory signal has already taken place, such that the Feature Detection Layer represents features in a subsymbolic but distal, rather abstract way (Heider, 1959 [1926]; Hommel et al., 2001). The patterns of activation in the input layer are categorized by the Object Module, by mapping different input patterns onto different nodes in the module.

Following categorization by the Object Module, information is propagated, through several layers of internodes (see Fig. 2), to a next layer, called *Memory Buffer*. The Memory Buffer is similar to the Conceptual Short Term Memory (Potter, 1993), retaining the most recently activated concepts. As a result, a pattern of activation arises in the Memory Buffer, representing the co-occurrence of objects across a stretch of time. This pattern is used as

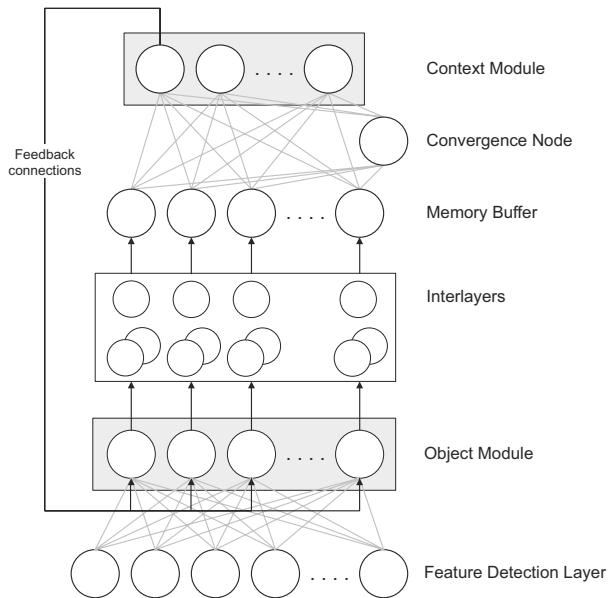


Fig. 1. CONCAT architecture.

input for the next level, the Context Module. Importantly, input patterns are not presented in completely random order, but they appear in coherent clusters of co-occurring patterns, which can be regarded as different contexts. As a result of this clustering, the activation patterns in the Memory Buffer show some regularities, which are picked up and categorized by the Context Module. In addition to this feedforward stream of information, feedback connections provide top-down input from the Context Module to the Object Module. These feedback connections are modified through a form of novelty-dependent competitive Hebbian learning (for details, see below). Due to the feedback connections, objects are effectively represented not only by bottom-up feature information but also by top-down context information.

2.2. Processes

The CONCAT model learns to categorize objects that are distributed over a number of different contexts. Therefore, input patterns are not presented in completely random order, but they are grouped within clusters of contextually related patterns. As a metaphor, think of someone walking through a house and opening doors to different rooms in that house. When looking into a specific room, attention shifts from one object to the other, in order to recognize and categorize those objects as, for example, a faucet, a toaster, an oven, and a fridge. Once a sufficient number of objects have been recognized, the room (i.e., context) can be reliably categorized as a kitchen. Had one seen a faucet, shower, towels, and a bath instead, the room would have been categorized as a bathroom. Thus, the collection of objects determines what kind of room (i.e., context) one is looking at. In order to simulate this procedure,

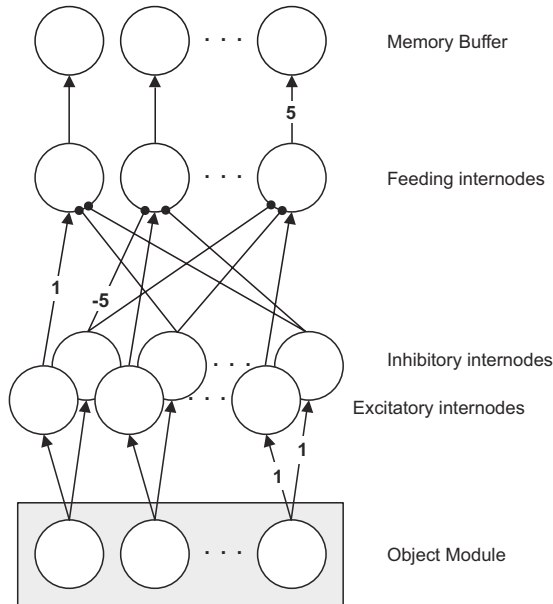


Fig. 2. Wiring of internodes.

several contexts are created. A context is defined as a collection of various input patterns (objects). Because input patterns may occur in multiple different contexts (just like some objects can be found in various different rooms), contexts are partially overlapping.

In order to learn the different contexts, the model experiences multiple contextual episodes (cf., opening the various doors in a house). During a contextual episode (see Fig. 3), a context is selected and all patterns belonging to the selected context are presented sequentially, in random order. A pattern is presented on the Feature Detection Layer, and categorized by the Object Module based on the bottom-up feature information. Activation is forwarded, through several layers of internodes, to the Memory Buffer. Although activation spreads through the network at every iteration, due to the wiring of the internodes (explained in more detail below) activation does not reach the Memory Buffer until the Object Module has reached convergence (i.e., the pattern has been categorized). In other words, the Memory Buffer only represents concepts that have been fully categorized by the Object Module. After a fixed number of iterations, the presentation of the input pattern is terminated. Before presenting a new input pattern from the same context, there is a delay during which the model does not receive any input pattern. This delay is implemented in order to let the activations of the Object Module decay to zero. Since the Memory Buffer has a much slower decay rate than the Object Module, the active concept is retained in the Memory Buffer throughout the delay that resets the Object Module. After the delay, the next pattern of the current context is presented and categorized in a similar way as the first pattern. Again, activation is forwarded to the Memory Buffer, where the previously activated node is still somewhat active. As a result, a pattern of activation gradually arises in the Memory Buffer, reflecting the most recently categorized concepts. Once the total amount of activation in the

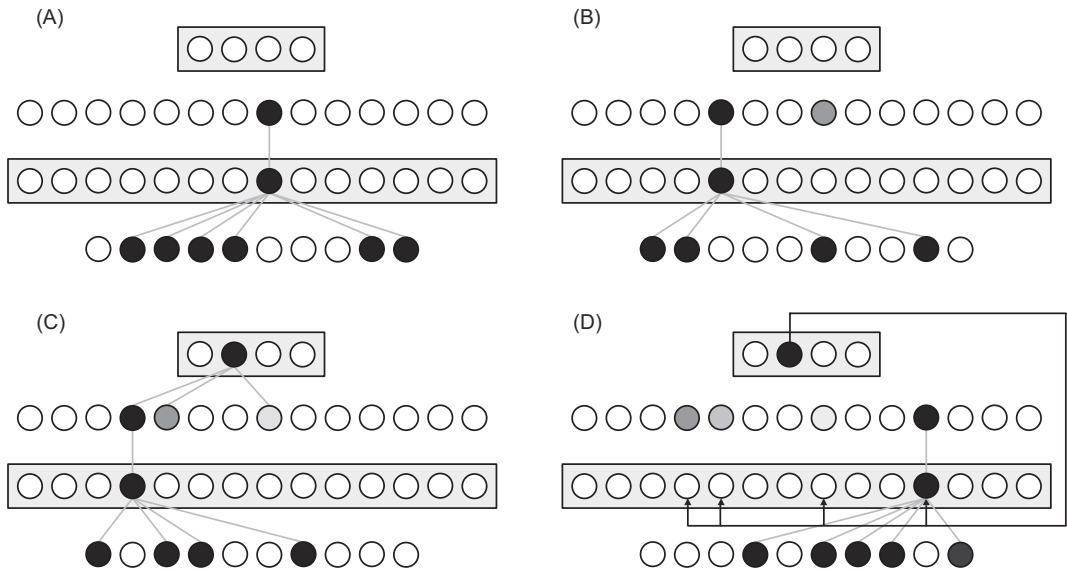


Fig. 3. Snapshots of activation propagation in the model. The figure only shows the Feature Detection Layer, Object Module, Memory Buffer, and Context Module. Four different input patterns belonging to the same context are presented. In Panel A, the first input pattern is presented, categorized by the Object Module and stored in the Memory Buffer. In Panel B, the second input pattern is categorized and stored in the Memory Buffer. In Panel C, the third input pattern is categorized and stored in the Memory Buffer. The total activation of the Memory Buffer now exceeds a certain value, resulting in activation of the Context Module, which categorizes the pattern in the Memory Buffer. In Panel D, the activated context node sends top-down (voltage-dependent) input to all associated concept nodes. This facilitates the categorization of a new pattern belonging to the same context.

Memory Buffer reaches a certain level, the pattern of activation is categorized by the Context Module.

From the Context Module, activation is propagated down to the Object Module through modifiable feedback connections. The strength of these connections is adapted through Hebbian-like learning, such that connections get stronger when the sending context node and the receiving object node are concurrently active, and they get weaker when one node is active while the other one is inactive. Thus, the strength of a feedback connection reflects how often a particular object has occurred in a given context. As a result, the winning context node sends top-down activation to objects proportional to their occurrence within this context. As will be demonstrated by several simulations (see below), this top-down context activation may influence (and often improve) the categorization of objects, especially when the bottom-up featural input is ambiguous or degraded. In the case of an ambiguous input pattern, multiple object nodes may get activated, resulting in a competition between those nodes. Top-down activation from the active context selectively increases the activation of the node representing the object that is most likely given the current context. This node, representing the “expected” object, therefore has a higher chance of winning the competition. A similar process takes place when the input is degraded. In this case, the bottom-up

input may be too weak to activate the correct object node to a sufficient degree to let it win the competition. The activation of this object node can be enhanced by a top-down signal from the current context. In other words, contextual feedback can make object nodes more sensitive to patterns belonging to the currently active context. This characteristic of the model is in accordance with several experimental findings that objects are more accurately recognized when they are presented within a relevant context than in isolation or in an irrelevant context (e.g., Davenport & Potter, 2004; Özcan & van Egmond, 2009).

Once all patterns of a context have been presented, the contextual episode is finished with an empty delay, in order to let all activations in the network decay to zero. Of course, in real life, episodes are not separated by a delay. Instead, information about the world enters our senses continuously. According to the Event Segmentation Theory (Zacks & Swallow, 2007), people parse this continuous input into discrete events.³ Simulations with a computational model of event segmentation have demonstrated that event boundaries can be reliably detected in a continuous stream of input (Reynolds, Zacks, & Braver, 2007). In CONCAT, this aspect was bypassed by presenting discrete episodes that were separated by a delay. However, the work of Zacks and colleagues (Zacks, 2004; Reynolds et al., 2007) shows that it is not unreasonable to assume that continuous input is parsed into discrete events, and that event boundaries can be used to reset the activations in the model.

2.3. Node activation rule

The input e_i to each network node i may be either excitatory (positive) or inhibitory (negative). The activation of a generic CONCAT node i at the iteration $(t + 1)$, $a_i(t + 1)$, is described by the following rule:

$$a_i(t + 1) = (1 - k) a_i(t) + \frac{e_i}{1 + e_i} [1 - (1 - k) a_i(t)] \quad e_i \geq 0 \quad (1a)$$

$$= (1 - k) a_i(t) + \frac{e_i}{1 + e_i} (1 - k) a_i(t) \quad e_i < 0 \quad (1b)$$

where $e_i = \sum_j w_{ij} a_j(t)$

k is an activation decay term, w_{ij} denotes the connection weight from node j to node i . This activation rule ‘‘squashes’’ the input to a number between 0 and 1, and ensures that the increase in activation due to input excitation, or the decrease in activation due to input inhibition, diminishes as the activation approaches the maximum or the minimum, respectively (asymptotic approach).

2.4. Internal structure of object and context modules

The Object Module and Context Module are so-called CALM modules (Murre et al., 1992), consisting of different types of nodes that are connected to each other (see Fig. 4) by fixed weights. The weight values were taken from Murre et al. (1992) and are provided in Table A1. We would like to refer the reader to the original paper of Murre et al. for the

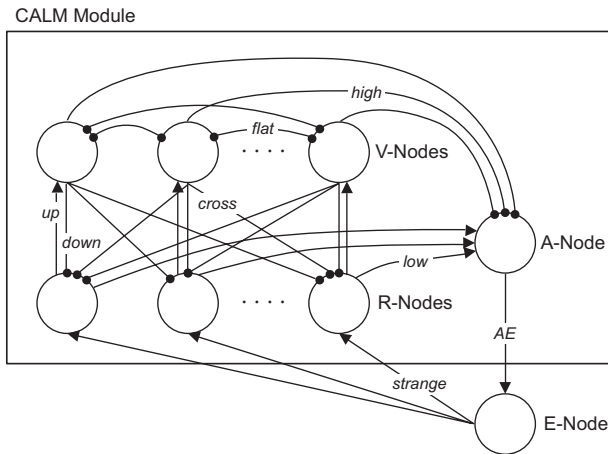


Fig. 4. CALM Module.

guiding principles underlying the model's structure and its parameters. "Representational-nodes" (furthermore called R-nodes) are the only nodes that receive external input. In order to create competition between R-nodes, each R-node activates a matched "Veto-node" (V-node), which inhibits all other nodes in the module. In particular, V-nodes are characterized by mutual inhibition; they strongly inhibit unmatched R-nodes, but they inhibit their matched R-node only weakly. The modules also contain an "Arousal-node" (A-node), which is sensitive to the novelty of the input pattern. In a sense, activation of the A-node can be regarded as a representation of the arousal response, which may result in an orientation reaction (e.g., Sokolov, 1963). Given the specific wiring pattern of the connections in a CALM module (see Fig. 4), the activation of the A-node is a positive function of the amount of competition in a module. Since competition is higher when a novel pattern is presented than when a familiar pattern is presented, the A-node activation essentially reflects the novelty of a given input pattern. Finally, an "External-node" (E-node), which might correspond to an arousal center external to the module (e.g., a subcortical center in the brain), receives input from the A-node of the module and sends random activation pulses to all R-nodes in the same module. These random pulses are uniformly distributed over the range $[0, a_E(t)]$, where $a_E(t)$ stands for the activation of the E-node at the iteration t .

2.5. Learning rule

Full connectivity is assumed from node layers projecting onto the Object Module (from the Feature Detection Layer and the Context Module) and the Context Module (from the Memory Buffer). The weights of these connections (interweights) are modifiable in the range $[0, K]$. The learning rule used in CALM networks (Murre et al., 1992) is a variation of Grossberg's (1976) learning rule, which is related to the Hebb rule (Hebb, 1949). The learning rule is expressed by the following formula:

$$\Delta w_{ij} = \mu_t a_i \left[(K - w_{ij}(t))a_j - Lw_{ij}(t) \sum_{f \neq j} w_{if}(t)a_f \right] \tag{2}$$

where i, j , and f are R-nodes, L and K are positive constants. The first term within the brackets represents weight increases, whereas the second term determines the background-dependent weight decrease. The symbol μ_t denotes the variable learning rate at iteration t , which is computed as follows:

$$\mu_t = d + w_{\mu E} a_E \tag{3}$$

where d is a constant and a_E is the activation of the E-node. The learning rate is thus dependent on the activation of the E-node, which is more active when a novel pattern is presented than when an old pattern is presented (for a more extensive explanation of the underlying process, see Murre et al., 1992). As a result, the learning rate of both the Object Module and the Context Module are dependent on the novelty of the pattern or context, respectively. Fig. 5 displays the learning rate of both modules during multiple presentations of the same input pattern. During the first presentation of a pattern or context, the learning rate is quite high, but it quickly drops to a lower level, indicating a shift from elaboration learning to activation learning.

To bind interweights in the range $[0, K]$, the following condition is included:

$$w_{ij}(t + 1) = \max\{\min[w_{ij}(t) + \Delta w_{ij}(t + 1), K], 0\} \tag{4}$$

The learning parameters used in the simulations are specified in Tables A2 and A3.

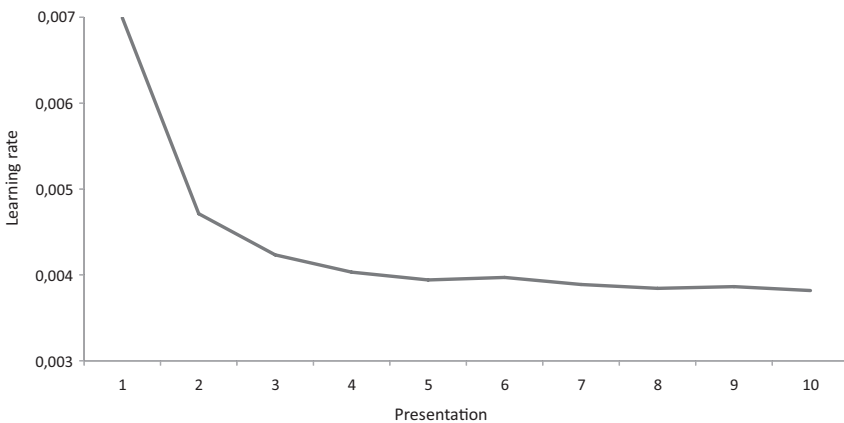


Fig. 5. Learning rate (μ) of the Object Module during subsequent presentations of the same input pattern.

2.6. Connection chain from the object module to the context module

As shown in Fig. 2, a sequence of intermediate nodes (internodes) is placed between the Object Module and the Memory Buffer, which in turn sends output to the Context Module. These internodes ensure that nodes in the Memory Buffer do not get activated until convergence has been reached in the Object Module. In other words, the Memory Buffer only retains objects that have been categorized by the Object Module. To this end, the R-nodes of the Object Module send excitatory input to a matched ‘‘Excitatory-internode’’ (E-internode) and to a matched ‘‘Inhibitory-internode’’ (I-internode). In turn, the matched E-internode forwards excitatory input to a matched ‘‘Feeding-internode’’ (F-internode), whereas the matched Inhibitory-internode forwards inhibitory input to all other unmatched F-internodes. Each F-internode, matched backward to an Object R-node, sends activation to a matched node in the Memory Buffer.

The excitatory input to the R-nodes of the Context Module includes input from individual Memory Buffer nodes, multiplied by modifiable interweights, and from a ‘‘Convergence node’’ (C-node), the excitatory input of which is described by a sigmoidal function of the sum of the activations of all Memory Buffer nodes, as follows:

$$input_c(x) = \frac{1}{1 + \exp(-\beta(x - \theta))} \quad (5)$$

As the Context Module R-nodes receive a fixed amount of tonic inhibition, *Inht*, the input from the C-node unselectively enables activation of the R-nodes in the Context Module, which becomes selective with the differential modification of the interweights from Memory Buffer nodes to Context Nodes. This enabling mechanism plays a functional role to make activation of the Context Module dependent on the number of active nodes in the Memory Buffer. The weight values and equation parameters of the model are specified in Tables A4 and A5.

3. Simulations

3.1. Simulation I: Simultaneous categorization of objects and contexts

CONCAT is not the first model that displays hierarchical categorization and the role of top-down feedback on categorization. The most well-known example is probably McClelland and Rumelhart’s (1981) interactive activation model. Although applied to a different domain (visual word and letter recognition), their model is based on similar assumptions as our model (see also Kintsch, 1988). First, they assume that perceptual information is processed at different levels, ‘‘each concerned with forming a representation of the input at a different level of abstraction’’ (p. 377). They distinguish three levels, representing visual features, letters, and words, which could be aligned with the three levels of

CONCAT, representing sensory-motor features, objects, and contexts. Furthermore, they assume that categorization at the middle level (the letter level) is the result of the interaction between bottom-up information from the feature level and top-down information from the word level, which is highly similar to the assumption of interactivity underlying CONCAT. However, an important difference is that in the interactive activation model the connection strengths between features and letters and those between letters and words are all predefined. This is not the case in CONCAT, where the connection values between the different levels are gradually acquired over the course of learning. In other words, the model learns from experience how features are distributed over objects and how objects are distributed over contexts.

Our model also bears some similarity to an other hierarchical word-letter recognition model based on the CALM architecture, developed by Murre et al. (1992). In that model, as in CONCAT, the connections between the different levels are gradually learned during the training stage. However, in contrast to CONCAT, the model of Murre et al. first has to learn the lower-level categories (letters) separately before it can process them together to perform the higher-level categorization (words). In the domain of object categorization, such a procedure is not very plausible. Objects are never encountered in isolation, but they are always surrounded by other objects and embedded in a larger context. Therefore, it is more likely that objects and contexts are learned simultaneously. This aspect was addressed in Simulation I.

The goals of Simulation I are two-fold. First, we aim to demonstrate that CONCAT can detect in its input regularities at multiple levels of abstractness and categorize these regularities as objects and contexts. Moreover, we aim to show that it can learn both levels of representations *simultaneously*. These aspects provide significant advantages over both the interactive activation model (McClelland & Rumelhart, 1981) and the model of Murre et al. (1992). The second goal of Simulation I is to investigate how the model's performance is influenced by the degree of overlap between the input patterns. The model's input was represented as a distributed binary pattern over 12 feature detection nodes. Two different sets of input patterns were created. In the *Low-overlap* condition, sparse patterns were used in which only two out of twelve nodes were "on." In the *High-overlap* condition, dense patterns were used in which, on average, six out of twelve nodes were "on." We expected that categorization performance would be better in the Low-overlap condition than in the High-overlap condition. The model was presented with 16 different input patterns, which were clustered across four contexts. Half of the concepts were presented in one context only; the other half of the concepts appeared in two different contexts. As a result, each context contained six different concepts, and contexts were partially overlapping.

The simulation procedure consisted of 140 episodes. The first 80 episodes of the simulation were regarded as training episodes. The last 60 episodes were regarded as test episodes during which the performance of the model was assessed. It is important to note that learning was not disabled during the test phase. However, since the learning rate is strongly dependent on the novelty of the pattern and no new patterns were presented during the test phase, learning was practically negligible. During each episode, a context was selected semi-randomly; we ensured that all contexts were presented equally often, both during the

Table 1
Example of output table

Input Pattern	No. of Presentations	Concept Node															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	15									15							
1	15				14			1									
2	15					15											
3	15										15						
4	15						15										
5	15							15									
6	15		2							13							
7	15		15														
8	15			14					1								
9	15														15		
10	15	15															
11	15													15			
12	15												15				
13	15															15	
14	15			1		13							2				
15	15					3							12				

training phase and during the test phase. Once a context was selected, each of the patterns contained in that context were presented in a random order. Each pattern was presented for a fixed number of iterations.

Activation was propagated through the model as described above. The specific parameters of the simulations are given in Table A6. During the test phase, the model's performance was assessed in the following way. After each pattern presentation, if the Object Module had reached convergence, the winning object node was registered. Similarly, at the end of each context presentation, the winning context node was recorded. After completion of the simulation, this information was aggregated into two different tables. One table displayed the number of times that each pattern converged on the different object nodes (see Table 1 for an example). A similar table was created for the contexts, displaying the number of times that each context converged on the different context nodes. The node that most often represented a particular input pattern was labeled as the "winner" for that pattern. Orthogonally, the pattern that was most often represented by a particular node was labeled as the winning pattern for that node. This information was used to compute two measures of performance. The *Convergence Index* was used as a measure of how well a particular input pattern was categorized by a single node. It was defined as the number of times that the particular pattern was mapped onto its winning node, divided by the total number of presentations of that pattern. This resulted in a value between 0 and 1, where 0 meant that the input pattern never reached convergence onto any node, and 1 meant optimal convergence onto a single node. Orthogonally, the *Separation Index* was used as a measure of how well a particular node came to represent a single pattern. The Separation Index was computed as the number of times that a particular node was converged on by its winning pattern, divided by

Table 2
Results of Simulation I

Input	Module	Convergence Index	Separation Index
Low overlap	Concept Module	0.98	0.96
	Context Module	0.94	0.95
High overlap	Concept Module	0.88	0.85
	Context Module	0.83	0.82

Note. Convergence Index and Separation Index of the Concept Module and Context Module under conditions of low and high pattern overlap.

the total number of times that it reached convergence. This resulted in a value between 0 and 1, where 0 meant that the node did not represent any pattern, and 1 meant that the node only represented its winning pattern. In addition to these measures, convergence times were logged for each pattern presentation. Convergence time was defined as the number of iterations needed by the Object Module to reach convergence.

The simulation procedure was replicated ten times for each of the two different input sets. This was done in order to get an indication of the variance in performance as a result of different random settings used in each of the replications. Convergence in the Object Module was influenced by the amount of overlap between the input patterns. In the Low-overlap condition, the Object Module needed on average 11.3 iterations to reach convergence. In the High-overlap condition, an average of 16.6 iterations was needed to reach convergence. Table 2 shows the Convergence Index and Separation Index of the Object Module and the Context Module. In both conditions of pattern overlap, the model performs quite well, resulting in high Convergence and Separation Indexes (all above 0.80). Although performance is better in the Low-overlap condition, the model still performs reasonably well in the High-overlap condition.

These results show that the model can learn to categorize concepts and contexts simultaneously. This is an important and novel finding. Previous simulations with a hierarchical CALM network have shown that it is possible to perform hierarchical categorization (Murre et al., 1992, 1992). In those simulations, however, the model first learned the lower-level categories separately, before learning the higher-level categories. In contrast, for CONCAT it is not necessary to learn object representations separately before learning them within their contexts. In other words, the current simulation demonstrates that it is possible to learn categories at both levels at the same time.

3.2. Simulation II: Influence of context feedback on categorization of ambiguous patterns

The aim of Simulation II was to investigate the role of context feedback on the categorization of ambiguous input patterns. Oliva and Torralba (2007) argued that the identification of ambiguous objects is constrained by the scene context. Similar effects may occur in language processing. For example, Zeelenberg et al. (2003) showed that semantic priming of ambiguous words (e.g., ORGAN) is context-dependent. During the first phase of their experiment, which served as an incidental learning task, participants performed a

Table 3

Example of overlapping patterns and ambiguous test pattern presented in Simulation II

Concept	Distributed Feature Pattern												Context
1	1	0	1	1	1	0	0	1	0	1	1	0	A
2	1	0	1	0	1	0	0	1	0	1	1	1	B
Test	1	0	1	0.5	1	0	0	1	0	1	1	0.5	A/B

word-completion task. Ambiguous words had to be completed in a sentence that biased one of its two possible meanings. For example, for the word ORGAN the sentence was either ‘‘The liver is an important o.... of the human body,’’ or ‘‘Kevin played hymns on the o....’’ In the second phase, participants performed a word association task. They received a cue word and had to write down the first related word that came to mind. Cues for the ambiguous words were related to one of the two possible meanings. For the word ORGAN the cue was either *liver* or *piano*. Ambiguous words were more likely to be produced if the cue (e.g., piano) was congruent with the sentence context of the first phase of the experiment (‘‘Kevin played hymns on the o....’’). This shows that context information strongly biases the interpretation of ambiguous words.

The influence of context on the interpretation of ambiguous stimuli was studied in Simulation II. The model learned to categorize two highly overlapping input patterns, among other, less overlapping, patterns. The highly overlapping patterns differed by only two nodes (out of a total of 12 nodes). During the training stage, the two overlapping patterns were presented in two different contexts. During the test stage, instead of the original patterns, an ambiguous pattern was presented that equally deviated from each of the two patterns (see Table 3 for an example). This input was presented to two different versions of the model. In the *Feedback-present* simulation condition, feedback input was allowed to propagate from the Context Module to the Object Module. This feedback input could facilitate categorization of ambiguous input patterns in the following way: Since the ambiguous pattern is equally similar to both original patterns, it will activate the two concept nodes representing those patterns to a similar degree. The object node that is most likely within the given context will receive some extra top-down activation from the active context node, which increases the probability that this node will win the competition. As a result, the ambiguous pattern will be categorized as either one of the original patterns, depending on the context in which it is presented. In the *Feedback-absent* simulation condition, no feedback was propagated from the Context Module down to the Object Module. Thus, context information could not be used to bias categorization toward the most likely object.

In this simulation, each context consisted of four objects. During the training stage, a contextual episode consisted of the random presentation of each of the patterns belonging to that context in random order, as was the case in Simulation I. During the test stage, patterns were not presented completely randomly. Instead, the ambiguous pattern was always the last pattern to be presented within a contextual episode. This was done to ensure that the context representation was already active at the time of presentation of the ambiguous pattern, such

that the categorization of that pattern could be biased by top-down information flowing from the activated context. The test phase was repeated with the original input patterns instead of the degraded patterns, in order to verify that the original patterns were categorized correctly. The model's performance in the "ambiguous" test stage was compared to its performance in the "original" test stage, which resulted in a measure of categorization accuracy. As in Simulation I, the whole procedure was replicated ten times with both versions of the model. The results of Simulation II are presented in Fig. 6, which displays the categorization of the ambiguous pattern. As becomes clear from the figure, context feedback influenced the categorization of the ambiguous pattern in a consistent manner. In the Feedback-present condition, the ambiguous pattern was most likely to be categorized as the object that was congruent with the pattern in which it appeared. No such bias existed in the Feedback-absent condition, where the pattern was equally likely to be categorized as the context-related object or the context-unrelated object. Note that this simulation does not show that context feedback necessarily results in higher accuracy than when no feedback is present. If a contextually incongruent pattern is presented, it may erroneously be categorized as the object that is more likely in the current context. However, given that objects most often occur in a congruent context, the context bias will usually lead to better performance.

3.3. Simulation III: Influence of context feedback on categorization of degraded patterns

The role of the feedback connections was further investigated in Simulation III. The goal of this simulation was to investigate how context information influences the categorization of degraded input. In real-life object categorization, the bottom-up feature input is often incomplete, for example, when the object is partially occluded from view, or when it is dark or noisy. In those cases, the context may facilitate categorization by priming those objects

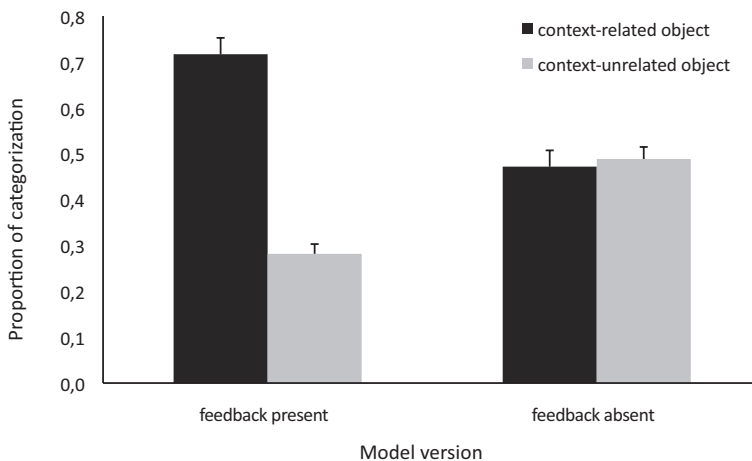


Fig. 6. Results of Simulation II. Object categorization of ambiguous input patterns by feedback-present or feedback-absent version of the model.

that are likely to occur. This has indeed been found in several studies. For example, Özcan and van Egmond (2009) showed that households producing sounds that were difficult to recognize when presented in isolation were recognized more accurately when congruent visual context information was presented. Similarly, Bar and Ullman (1996) showed that unclear images were recognized faster and more accurately in the presence of a contextually related image than in isolation.

During the test phase of this simulation, the model received a number of degraded input patterns, and categorization performance on these patterns was assessed. One of the patterns from each context was degraded by reducing all units with a value of 1 and increasing all units with a value of 0 by a fixed value, corresponding to the severity of degradation. Ranging from mild to severe degradation, patterns were degraded by a value of 0.2, 0.25, 0.3, 0.35, and 0.4, respectively. Similar to the procedure of Simulation II, the input patterns were presented to two versions of the model. In the *Feedback-present* condition, feedback was propagated from the Context Module down to the Object Module. In the *Feedback-absent* condition, no such feedback was propagated through the network. Since the bottom-up input from the Feature Detection Layer to the Object Module is degraded, it may not activate the correct concept node to a sufficient degree to let it win the competition. Top-down activation from the active context node may facilitate the categorization, by enhancing the object nodes that are most likely given the current context. Therefore, we expected that performance in the Feedback-present condition would be better than in the Feedback-absent condition.

The training phase was similar to that of Simulation II, except that there were no pairs of highly overlapping input patterns. As in Simulation II, each context included four objects, such that contexts were nonoverlapping. During the training stage, a contextual episode consisted of the random presentation of each of the object patterns belonging to that context. During the test stage, the degraded pattern was always presented last within a contextual episode. This was done to ensure that the Context Module had already categorized the context at the time of presentation of the degraded pattern, such that the categorization of that pattern could be biased by top-down information flowing from the active context node.

The results of Simulation III are displayed in Fig. 7. Fig. 7A shows the categorization accuracy of the Feedback-present and the Feedback-absent versions under the different degrees of degradation. As predicted, categorization performance was better in the Feedback-present than in the Feedback-absent condition, except when the pattern was only mildly (20%) degraded.

Fig. 7B shows the average number of iterations needed to reach convergence. Convergence times were faster in the Feedback-present version than in the Feedback-absent version. These findings corroborate the results of Simulation II, by demonstrating how context information can constrain object categorization. This is especially useful when the bottom-up input is ambiguous or degraded and therefore provides insufficient information to enable successful categorization of the pattern. When the input pattern is complete or only mildly degraded, there is no difference in performance between the Feedback-present and Feedback-absent version of the model.

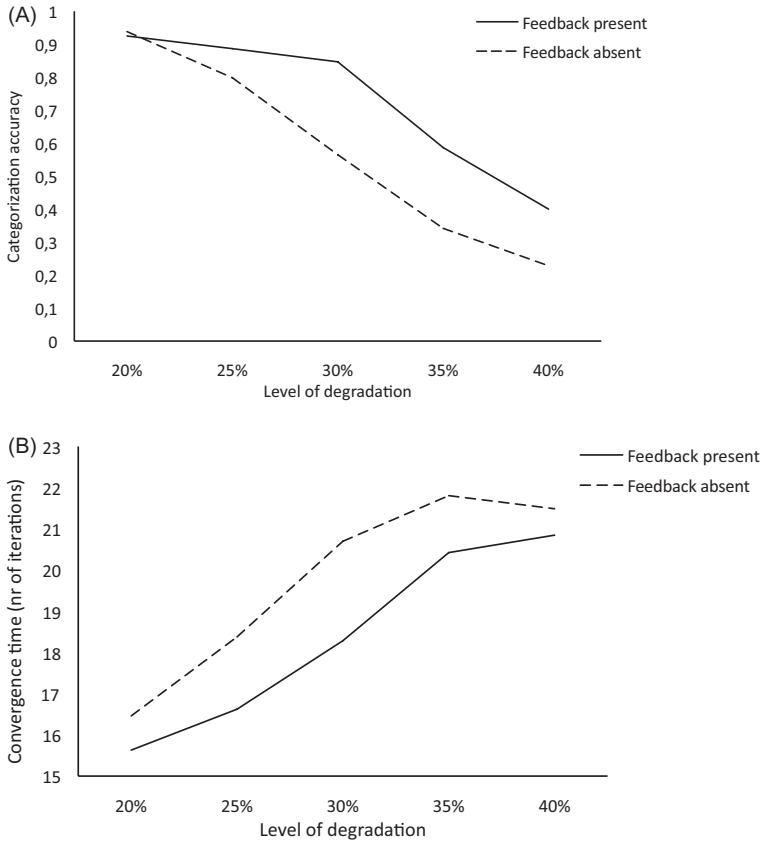


Fig. 7. Results of Simulation III. Average categorization accuracy (A) and convergence time (B) of degraded input patterns by the feedback-present version (continuous line) or feedback-absent version (dashed line) of the model.

3.4. Simulation IV: Categorization of objects in congruent and incongruent contexts

Several studies demonstrate that objects are recognized faster and more accurately when they are presented in a congruent context than when they are presented in an incongruent context (e.g., Davenport, 2007; Davenport & Potter, 2004; Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Palmer, 1975). This aspect of concept categorization was studied in Simulation IV. In the test phase of this simulation, complete and degraded input patterns were presented either in the same context as in the training phase (*congruent context*) or in a different context than in the training phase (*incongruent context*). We expected that patterns presented in a congruent context would be categorized more accurately than patterns presented in an incongruent context. The training phase was similar to Simulations II and III. During the test stage, the test pattern was either presented in a congruent context or in an incongruent context. As in Simulations II and III, test patterns were always presented last

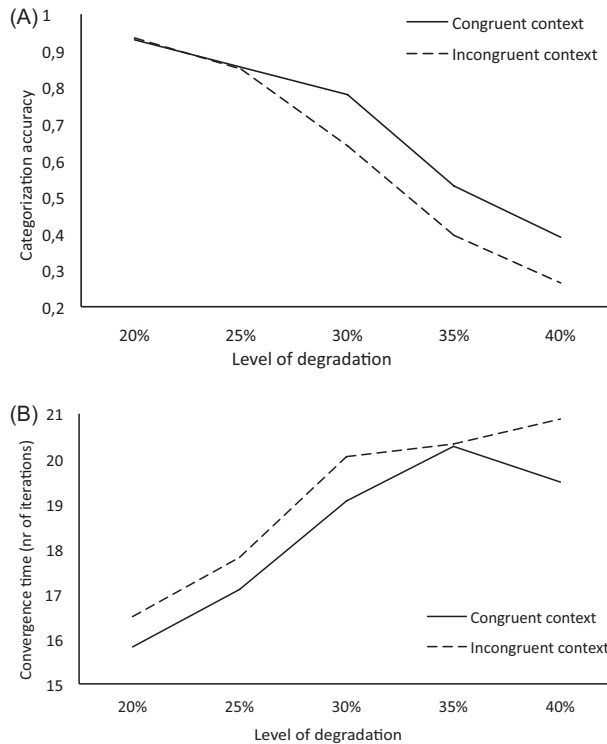


Fig. 8. Results of Simulation IV. Average categorization accuracy (A) and convergence time (B) in a congruent context (continuous line) or incongruent context (dashed line).

within a contextual episode, to ensure that the context had been categorized before the test pattern was presented.

The results of Simulation IV are displayed in Fig. 8. Fig. 8A shows the categorization accuracy of the model in the congruent and incongruent context for the different degrees of degradation. Fig. 8B shows the convergence times in the congruent and incongruent context. As predicted, the model performed better in the congruent context than in the incongruent context. In addition, convergence times were slightly faster in the congruent context than in the incongruent context. These results are in line with behavioral data demonstrating that object recognition is faster and/or better when an object is presented in a congruent context than when it is presented in an incongruent context.

4. General discussion

In this paper, we argue that people create mental representations at multiple hierarchical levels. We specifically focus on object representations and context representations. We assume that both types of representation are acquired through experience, by extracting the statistical regularities that emerge during repeated experiences. Features that frequently

co-occur in time and space are likely to belong to the same object. At a coarser temporal and spatial scale, objects that frequently co-occur are likely to belong to the same context. Thus, while objects can be defined as patterns of co-occurring features, contexts can be defined as higher-order patterns of co-occurring objects. We furthermore propose that object representations and context representations are dynamically connected to each other. When an object representation is activated, activation will spread toward the contexts in which it frequently occurs, and reversely, when a context representation is activated, activation will spread toward the objects that are likely to occur in this context. As a result, cognitive processing can be tailored to the current situation. As Yeh and Barsalou argued, ‘‘because specific entities and events tend to occur in some situations more than others, capitalizing on these correlations constrains and thereby facilitates processing. Rather than having to search through everything in memory across all situations, the cognitive system focuses on the knowledge and skills relevant in the current situation. Knowing the current situation constrains the entities and events likely to occur. Conversely, knowing the current entities and events constrains the situation likely to be unfolding’’ (Yeh & Barsalou, 2006, p. 350).

Consistent with this reasoning, the results of our simulations show that context information can indeed play an important role in guiding and constraining conceptual processing. Our findings are in line with many studies across different domains of cognitive science (perception, language comprehension, memory, concept representation), demonstrating a variety of context effects (e.g., Barsalou, 1982; Chaigneau et al., 2009; Davenport & Potter, 2004; Godden & Baddeley, 1975; Mandler & Parker, 1976; Miller & Gazzaniga, 1998; Özcan & van Egmond, 2009; Zeelenberg et al., 2003). Furthermore, in computer vision, object recognition can be improved by employing context information based on the interaction between objects or global scene statistics (Galleguillos & Belongie, 2010).

An interesting and novel aspect of our model is that context information is essentially derived from the bottom-up input. The different levels of the model detect structure (correlational patterns) in this input at different levels of abstractness (or grain size). This distinguishes CONCAT from other models (e.g., Meeter, Talamini, & Murre, 2002; Rudy & O’Reilly, 2001), in which context information typically comes from an external source.

We do not claim that CONCAT is an exhaustive model of concept representation. Some empirical findings are inconsistent with the model. For example, it has been shown that visual scenes (contexts) can be recognized at a glance (e.g., Oliva, 2005), and then facilitate recognition of objects within (e.g., Bar, 2004; Torralba, 2003). This suggests that context is not only activated indirectly through its objects but also directly through rapid processing of its global features and coarse spatial information (Oliva & Torralba, 2001; Schyns & Oliva, 1994). Furthermore, although CONCAT is composed of CALM modules which have some biologically plausible properties (e.g., modularity, unsupervised learning and novelty-dependent learning), we are hesitant to connect the elements of CONCAT to specific brain areas or neurological processes. The model is based on assumptions about the hierarchical structure of the conceptual system and the way in which representations of different levels may interact, and it should be regarded as a feasibility test of these assumptions.

The idea that the conceptual system is composed of multiple hierarchically organized modules bears a similarity to Damasio’s theory of Convergence Divergence Zones

(Damasio, 1989; Meyer & Damasio, 2009). Convergence divergence zones (CDZs) are brain areas that function as association areas by registering the temporal co-occurrences between patterns of activation in lower-level areas. During interaction with an object, patterns of activation arise across different sensorimotor areas, coding for various aspects of the interaction. The temporally coincident activity at separate sites is bound by a CDZ. In order to recall a particular event, the CDZ (partially) reconstructs the original pattern of activation in the sensorimotor sites through divergent back projections. Damasio argues that multiple convergence zones can be organized in a hierarchical manner, each representing input at different levels of abstractness: “There are convergence zones of different orders; for example, those that bind features into entities, and those that bind entities into events or sets of events, but all register combinations of components in terms of coincidence or sequence, in space and time” (Damasio, 1989, p. 26). Although CONCAT currently encompasses only two levels, we assume that it is possible to extend its logic to build a network composed of multiple layers that categorize patterns occurring at various spatial and temporal scales. In addition, we assume that it is possible to implement feedback connections between those layers, similar to the divergent pathways proposed by the CDZ theory.

Another parallel can be drawn to a recent theory of word meaning, proposed by Andrews et al. (2009). They argue that word meaning is based on two sources of information. The first source is information about the sensory-motor properties of the word’s referent, as has been proposed in many theories of word meaning (e.g., Collins & Quillian, 1969; McRae, de Sa, & Seidenberg, 1997; Rips, Shoben, & Smith, 1973). The second source is information about the lexical contexts in which a word occurs, as proposed by theories such as Latent Semantic Analysis (Landauer & Dumais, 1997) and Hyperspace Analogue to Language (Lund & Burgess, 1996). The model of Andrews et al. (2009) integrates both sources of information (for similar work, see Steyvers, 2010). The data generated by the combined model corresponded better to human data of semantic representations than data generated by the separate sources. Thus, this study showed that experiential and distributional information can be combined to yield rich and realistic representations, similar to the findings of the present paper.

Andrews et al. (2009) define distributional data as information about the distribution of words across linguistic contexts, and they call this type of information “intra-linguistic.” We would like to argue that this information might not be purely linguistic, but that it could indirectly reflect the actual distribution of the words’ referents across real-world situations. After all, language refers to objects and situations in the world. Words might co-occur in language because their referents co-occur in the real world. For example, distributional theories of word meaning consider the words “*cat*” and “*dog*” as similar because they often occur in similar sentences. However, the reason that they occur in similar sentences might be because these sentences typically describe real-world situations in which cats and dogs occur. In the real world, cats and dogs are often found in similar contexts (e.g., in and around the house). As a result, the sentences describing these contexts will be similar. Indeed, a recent study by Louwerse and Zwaan (2009) showed that the spatial distance between pairs of U.S. cities can be derived from their distributions across texts. In the words of Louwerse and Zwaan (2009), “cities that are located together are debated together.”

To sum up, while we agree with Andrews et al. (2009) that distributional data form an important source of information about semantic or conceptual representations, we do not agree that this information is completely linguistic. Instead, the distribution of a word across linguistic contexts might correspond strongly to the distribution of the word's referent across real-world contexts (for a similar argument, see Zwaan & Madden, 2005).

Summarizing, we have stressed the importance of context information in conceptual processing. In addition, we have demonstrated how context information can be extracted from experience and categorized into context representations. These representations, through their connections with associated object representations, can bias and constrain the categorization of objects, which is especially useful when the bottom-up input is degraded or ambiguous. Including context information makes the conceptual system more adaptive and efficient. Another important consequence of such mechanism is that it allows abstract concepts, which have no perceptual features themselves, to be grounded in contextual experiences. Indeed, Barsalou and Wiemer-Hastings (2005) demonstrated that abstract concepts are largely based on situational information. They argue that it is impossible to represent an abstract concept like TRUTH without providing information about the situation in which this concept is set. Furthermore, although concrete concepts are typically processed better and faster than abstract concepts, these benefits disappear when abstract concepts are presented within a relevant situation (e.g., Schwanenflugel, Harnishfeger, & Stowe, 1988; Schwanenflugel & Stowe, 1989). Thus, context information appears to be highly relevant for the representation of abstract concepts.

We live in a world full of regularities. Not only do we repeatedly encounter similar things, but we also experience similar situations and find ourselves in similar environments over and over again. Our conceptual system is specialized in extracting these regularities from experience, in order to form mental representations of objects, events, and situations (e.g., Chun & Jiang, 1999; Fiser & Aslin, 2001). This results in a rich hierarchically organized knowledge structure that may represent objects and contexts, as well as the connections between them. Capitalizing on these connections facilitates processing in many ways. It enables one to select the representations that are relevant given the current context, to know which objects to expect in a given context, where to look for those things, and how to interact with them. In other words, it enables one to interact with the environment in a sensible and efficient manner.

Notes

1. Throughout this paper, angle brackets will be used to describe features (e.g., <round>), uppercase font will be used to describe concepts (e.g., BALL), and uppercase italic font will be used to describe contexts (e.g., *KITCHEN*).
2. The term *distributional* data is borrowed from Andrews, Vigliocco, and Vinson (2009). In their paper, however, the term had a slightly different meaning. The authors referred to experiential (feature-based) and distributional information as two contributors of word meaning. In their definition, distributional information refers to the

linguistic contexts (sentences, paragraphs) in which a word occurs. In the current paper, distributional information refers to the actual contexts in which a concept occurs.

3. The Event Segmentation Theory (Zacks, 2007) supposes that the perceptual system not only perceives current input but also predicts what will be perceived next. This prediction is based on the current input as well as on more stable information about the ongoing event, which is represented as an *event model*. Event boundaries are usually characterized by a relatively high *prediction error*, which occurs when the system's prediction is substantially different from the actual input. Whenever an event boundary is perceived, the event model is reset in order to represent the new event.

References

- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*, 463–498.
- Bar, M. (2004). Visual objects in context. *Nature Reviews: Neuroscience*, *5*, 617–629.
- Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*, 343–352.
- Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82–93.
- Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). New York: Cambridge University Press.
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.
- Chaigneau, S. E., Barsalou, L. W., & Zamani, M. (2009). Situational information contributes to object categorization and inference. *Acta Psychologica*, *130*, 81–94.
- Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365.
- Collins, A., & Quillian, M. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *12*, 240–247.
- Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*, 25–62.
- Davenport, J. L. (2007). Consistency effects between objects in scene processing. *Memory & Cognition*, *35*, 393–401.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*, 559–564.
- Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.
- Galleghillo, C., & Belongie, S. (2010). Context based object categorization: A critical survey. *Computer Vision and Image Understanding*, *114*, 712–722.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*, 325–331.
- Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior*, *23*, 553–568.
- Grossberg, S. (1976). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, *23*, 187–202.

- Hampton, J. A., Dubois, D., & Yeh, W. (2006). The effects of pragmatic context on classification in natural categories. *Memory & Cognition*, *34*, 1431–1443.
- Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley.
- Heider, F. (1959 [1926]). Thing and medium. *Psychological Issues, Monograph 3* (original work published 1926).
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–878.
- Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*, 1–18.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163–182.
- Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, *104*, 211–240.
- Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science*, *33*, 51–73.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation, and Computers*, *28*, 203–208.
- Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*, 38–48.
- Mandler, J. M., & Stein, N. (1974). Recall and recognition of pictures by children as a function of organization and distractor similarity. *Journal of Experimental Psychology*, *102*, 657–669.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, *88*, 375–407.
- McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. In B. H. Ross (Ed.), *Psychology of learning and motivation: Advances in research and theory*, Vol. 45 (pp. 41–86). San Diego, CA: Academic Press.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*, 547–559.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*, 99–130.
- Meeter, M., Talamini, L. M., & Murre, J. M. J. (2002). A computational approach to memory deficits in schizophrenia. *Neurocomputing*, *44*, 929–936.
- Meyer, K., & Damasio, A. (2009). Convergence and divergence in a neural architecture for recognition and memory. *Trends in Cognitive Science*, *32*, 376–382.
- Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia*, *36*, 513–520.
- Murphy, G. L. (2004). *The big book of concepts*. Cambridge, MA: The MIT Press.
- Murre, J. M. J., Phaf, H., & Wolters, G. (1992). CALM: Categorizing and learning module. *Neural Networks*, *5*, 55–82.
- Murre, J. M. J. (1992). *Learning and categorization in modular neural networks*. Hillsdale, NJ: Erlbaum.
- O'Connor, C. M., Cree, G. S., & McRae, K. (2009). Conceptual hierarchies in a flat attractor network: Dynamics of learning and computations. *Cognitive Science*, *33*, 665–708.
- Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *In the encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Science*, *11*, 520–527.
- Özcan, E., & van Egmond, R. (2009). The effect of visual context on the identification of ambiguous environmental sounds. *Acta Psychologica*, *131*, 110–119.

- Page, M. P. A. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, 23, 443–467.
- Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 3, 519–526.
- Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156–161.
- Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., & Sherman, J. (1986). Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General*, 115, 281–294.
- Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, 31, 613–643.
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1–20.
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rudy, J. W., & O'Reilly, R. C. (2001). Conjunctive representations, the hippocampus, and contextual fear conditioning. *Cognitive, Affective, and Behavioral Neuroscience*, 1, 66–82.
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, 27, 499–520.
- Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, 24, 232–252.
- Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, 24, 114–126.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, 5, 195–200.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214–241.
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. New York: Macmillan.
- Steyvers, M. (2010). Combining feature norms and text data with topic models. *Acta Psychologica*, 133, 234–342.
- Torralba, A. (2003). Contextual priming for object detection. *International Journal in Computer Vision*, 53, 169–191.
- Tversky, B., Zacks, J. M., & Martin, B. (2008). The structure of experience. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 436–464). Oxford, UK: Oxford University Press.
- Wolters, G., & Phaf, R. H. (1990). Implicit and explicit memory: Implications for the symbol-manipulation versus connectionism controversy. *Psychological Research*, 52, 137–144.
- Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology*, 119, 349–384.
- Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, 28, 979–1008.
- Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, 16, 80–84.
- Zeelenberg, R., Pecher, D., Shiffrin, R. M., & Raaijmakers, J. G. W. (2003). Semantic context effects and priming in word association. *Psychonomic Bulletin & Review*, 10, 653–660.
- Zwaan, R. A., & Madden, C. J. (2005). Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 224–245). Cambridge, UK: Cambridge University Press.

Appendix

Table A1

Weight values of connections within CALM Modules, as specified by Murre et al. (1992)

Weights	Description	Value
Up	Connects R-node to its matched V-node	0.5
Down	Connects V-node to its matched R-node	-1.2
Cross	Connects V-node to all nonmatched R-nodes	-10.0
Flat	Interconnects all V-nodes (no self-connections)	-1.0
Low	Connects R-node to A-node	0.4
High	Connects V-node to A-Node	-0.6
AE	Connects A-node to E-node	1.0
Strange	Connects E-node to R-nodes	0.5
Inter	Initial value of intermodular adaptive weights	0.6

Table A2

Equation parameters of Object Module

Globals	Description	Value
k	Decay of activation in an iteration	0.05
L	Learning competition between bottom-up connections	1.0
K	Maximum value of bottom-up weights	1.0
d	Base rate of learning	0.001
$W_{\mu E}$	Virtual weight between E-node and learning rate	0.015
C_h	High convergence threshold	0.001
C_l	Low convergence threshold	0.0001

Table A3

Equation parameters of Context Module

Globals	Description	Value
k	Decay of activation in an iteration	0.025
L	Learning competition between connections	1.0
K	Maximum value of learning weights	1.0
d	Base rate of learning	0.005
$W_{\mu E}$	Virtual weight between E-node and learning rate	0.025
C_h	High convergence threshold	0.001
C_l	Low convergence threshold	0.0001

Table A4
Weight values of connections between different layers of the model

Weights	Description	Value
R-I	Connects R-node in Object Module to matched Inhibitory internode	1
R-E	Connects R-node in Object Module to matched Excitatory internode	1
I-F	Connects Inhibitory internode to all nonmatched Feeding nodes	-5
E-F	Connects Excitatory internode to matched Feeding node	1
F-M	Connects Feeding node to matched Memory Buffer node	5
M-Cv	Connects Memory Buffer nodes to Convergence node	1
Cv-C	Connects Convergence node to R-nodes in Context Module	2
M-C	Initial value of connections between Memory Buffer nodes and R-nodes in Context Module	0.6
FB	Initial value of Feedback connections (from R-nodes in Context Module to R-nodes in Object Module)	0.6

Table A5
Equation parameters of other elements of the model

Parameter	Description	Value
k_{inter}	Decay of activation of internodes	0.1
k_{mem}	Decay of activation of nodes in Memory Buffer	0.001
β	Gain of input function of Convergence Node	-20
θ	Threshold of input function of Convergence Node (dependent on number of concepts per context)	1.5-3
Inht	Tonic inhibition to R-nodes in Context Module	-1.05
L_{fb}	Learning competition of feedback connections	1.5
K_{fb}	Maximum value of feedback weights	1.5
d_{fb}	Base rate of learning of feedback connections	0.005

Table A6
Simulation parameters

Description	Value
Number of nodes in Feature Detection Layer	12
Number of nodes in Object Module	16
Number of nodes in Context Module	4
Iterations per concept presentation	60
Number of context episodes during training	80
Number of context episodes during testing	20
Number of iterations during short reset interval	120
Number of iterations during long reset interval	3000