# A Neural Process Model of Structure Mapping Accounts for Children's Development of Analogical Mapping by Change in Inhibitory Control

**Minseok Kang (minseok.kang@ini.rub.de)**

**Daniel Sabinasz (daniel.sabinasz@ini.rub.de)**

**Gregor Schöner (gregor.schoener@ini.rub.de)**
Institute for Neural Computation, Ruhr-Universität Bochum
Universitätsstr. 150, D-44801 Bochum, Germany

## Abstract

We present a neural process model of visual analogical mapping that receives image inputs and responds by spatially selecting a matching object to a cued object. The relational structure of the base scene is stored in a way that specifies the arguments of each relation, allowing mappings based on structural correspondence to be represented as proposed by the structure-mapping theory. All the processes in the model emerge out of coupled integro-differential equations modeling neural population activation dynamics. The mapping can be influenced by both featural and relational similarities. The developmental shift in mappings in the presence of a featural distractor can be accounted for by manipulating how well the model can maintain attention to relevant feature/relation dimensions, consistent with a hypothesis suggesting inhibitory as a key factor explaining the shift.

**Keywords:** Analogy; Structure-mapping; Development; Inhibitory control; Computational modeling; Neural dynamics

## Introduction

Analogical mapping establishes the correspondence between objects and relations across two scenes based on the matching relation. According to the structure-mapping theory, analogical mapping is influenced stronger by relational similarity than featural similarity as mapping preserving the relational structure is preferred (Gentner, 2003). From the perspective of embodied cognition, analogical mapping plays an important role in abstract concept acquisition and comprehension by detecting similarities to concepts linked closer to sensorimotor surfaces (Lakoff & Johnson, 1980).

Studying developmental changes in mappings provides functional constraints to understanding the processes involved in analogical mapping. Many studies have shown that younger children fail to solve analogies as they are more likely to select an object that is similar in terms of their features rather than taking the relational structure into account (Gentner & Toupin, 1986; Markman & Gentner, 1993; Rattermann & Gentner, 1998). One possible explanation is that younger children are worse at inhibiting objects that are featurally similar to instead select an object that matches in terms of their relations (Richland, Morrison, & Holyoak, 2006). The intrusiveness of featural similarity suggests that analogical mapping is based on a general mechanism of similarity detection (Rattermann & Gentner, 1998).

Many computational models of analogical mapping have been proposed, and some of them generate a sequence of cognitive processes that are hypothesized to take place when solving analogies (Gentner & Forbus, 2011). The connectionist model by Hummel and Holyoak (1997) starts with a structured representation of two scenes and learns mappings between localist units based on the shared semantics. The model was constrained by the fact younger children are more likely to make mappings based on features. It does not include a visual attention system, and descriptions of visual scenes had to be hand-coded in activations of localist units. The model by Hesse, Sabinasz, and Schöner (2022) includes a visual attention system that takes real image input. However, the model was limited to processing only a single pairwise relation and did not respond spatially. It did not explicitly connect to existing literature on the development of analogical mapping.

This paper aims to demonstrate a neural process model of visual analogical mapping that can generate a sequence of cognitive processes leading from visual input to the spatial selection of a matching object. Dynamic Field Theory (DFT) is a suitable theoretical framework for building such a model (Schöner, Spencer, and DFT Rearch Group (2016)). DFT aims to explain how cognitive functions emerge from neural population activation dynamics modeled by coupled integro-differential equations. In DFT, general principles that hold across all areas of cognition can be found. A key principle is that decisions must be stabilized against fluctuating input. The underlying activation states are then attractors of the neural dynamical system. When the activation is above a threshold, local excitation stabilizes the activation, forming a localized *peak* along the feature dimension. Such decisions must be destabilized in order to generate sequences of mental states. This happens through bifurcations that are induced as input changes. For example, when the strength of a stimulus increases sufficiently, the system switches from a no-detection to a detection neural state by going through a *detection instability*. The activation variable time course makes different attractors appear and disappear, autonomously generating a sequence of cognitive processes.

Three key problems addressed by the model are highlighted. First, the relational structure of a scene has to be represented in a way that specifies i) which objects are involved in each relation and ii) which object is playing which role within a specific relation (Doumas & Hummel, 2005). In Figure 1, the description of the base scene should represent that it is the cat chasing the rat, not the dog, and that the rat is the one being chased, not the cat. Second, the model



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

receives real visual input and spatially selects a matching object. The model generates a sequence of neural states that describes the cognitive processes of perceiving the stimuli, figuring out the mapping, and indicating which objects are mapped onto each other. An explicit neural representation of the structure mapping enables spatially selecting the matching objects. Finally, the model captures the fact that younger children are more likely to select the featural distractor. Both the relational/featural similarities have an influence on its response, while younger children are assumed to have a weaker control of which kinds of similarities to use.

## Model

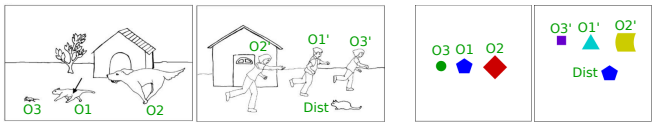


Figure 1: *left*) The example stimulus with a distractor. The figure is reprinted from Richland et al. (2006). *right*) The stimulus used for the present study. The numbering of objects is for illustration purposes only.

The model maps objects across two scenes and responds which object in the target scene corresponds to the cued object. A similar task had to be solved in the experiment by Richland et al. (2006), in which an arrow cued an object in the base scene.

We use stimuli defined over dimensions that can be directly grounded in terms of their perceptual features (Figure 1). The objects varied in color, shape, size, and spatial location. The relative size and the spatial relation between a pair of objects were treated as relations, while the colors and shapes of objects were treated as features of individual objects. Neutral items in the original stimuli were left out, as children were guided to focus only on critical objects. Figure 2 shows the sub-networks within the model and their interactions. The following processing phases emerge autonomously in the model.

First, the model **describes** the base scene and stores the relational structure in the conceptual description sub-network. The spatial selection sub-network sequentially selects an object to attend to. Each time an object is selected, its features are extracted in the feature processing sub-network, and its relations to one of the previously selected objects are extracted in the relational reasoning sub-network. Extracted features and relations are stored in the conceptual description sub-network.

Once the description of the base scene is completed, the model identifies which stored features and relations are critical for the decision to accept or reject the proposed mapping. Visual WM fields are inhibited to 'reset' them so the target scene can be processed. The visual input is changed to the target scene picture.

Next, the model **visually grounds** the description of the

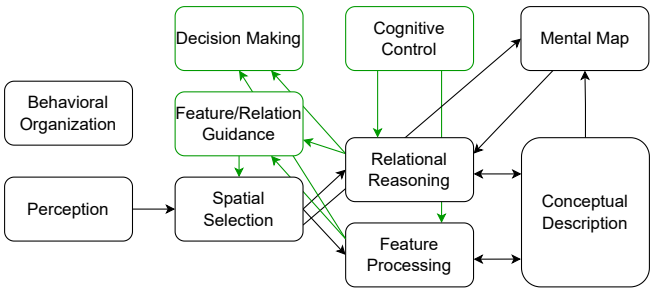


Figure 2: Abstract diagram of the model. Fields are classified into different functional sub-networks and depicted as boxes. Connections between different sub-networks are depicted as arrows. Black boxes are sub-networks needed for both the description and the grounding phase; green boxes are needed only from the grounding phase onward.

base scene in the target scene. Descriptions of objects are activated in the same order as the description order. Only a single object can be grounded at any given moment in time. Activated concepts in the conceptual description sub-network boost objects with matching features/relations in the guidance sub-network, which biases the spatial selection. The cognitive control sub-network controls whether the feature/relations can bias spatial selection.

After selecting a potentially matching object, the decision to **accept or reject** the mapping is made. Match/mismatch is detected in the decision making sub-network for all feature/relation dimensions, and both influence the decision. If the selected object does not yet provide enough evidence for the mapping, the architecture proceeds the grounding phase by **selecting another object** to check the relational match. If the selected object is accepted, the architecture **responds** by spatially selecting the object in the target scene that maps onto the cued object in the base scene. If the selected object is rejected, an **alternative mapping** is tried out. The grounding process starts from the start again, but the initial selection will be biased not to select the same object again, leading to a different mapping. Figure 3 shows all the neural fields and connections between them in the model. While they are divided into different sub-networks based on their functions, functions emerge out of connections between and within fields.

The **Perception / Spatial Selection / Feature Processing / Guidance** sub-networks provide the interface to sensory input and model visual attention. They resemble the visual search model by Grieben et al. (2020) in a simplified way. The spatial location and features of objects in the input image are conjunctively represented in three *space-feature maps*. The *spatial selection field* operates in a selective regime so that only one localized peak can form. Different inputs from the *saliency map*, the *inhibition-of-return* field, and the *guidance fields* sum up, and the location with the strongest input most likely form a peak. The *inhibition-of-return field* prevents the same location from being selected again. The *fea-*

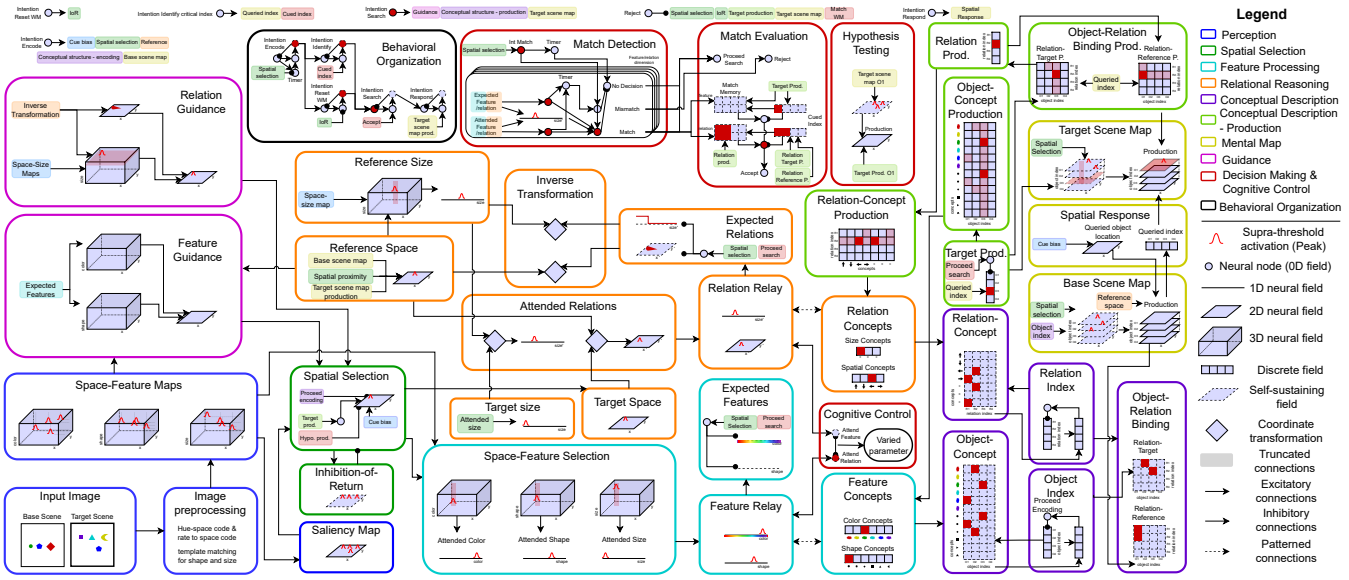


Figure 3: A detailed view of the model. Fields are grouped into different functional sub-networks by colored boxes. Input from some nodes is depicted externally on the top as truncated inputs.

*ture relay fields* have bi-directional patterned synaptic connections to the matching *feature concepts nodes* in the selective regime, expressing each concept's prototypical feature. The *guidance fields* bias the spatial selection to the location of the object whose features and relations match the feature and relations of the object in the base scene.

The **Relational Reasoning** sub-network describes the pairwise relations between two objects and uses the activated relation concepts to guide the spatial selection during the grounding phase. For each relation dimension (spatial/size), two distinct roles in a relation are represented by two separate fields. The *target fields* are filled by the object currently selected in the *spatial selection field*. The *reference space field* receives input from the *base scene map field* and the *target space field*, biasing the selection towards an object that was recently described and is spatially closed to the target object as the reference object. The *attended relation fields* represent the target object's values in shifted dimensions so that the reference object's values are represented in the middle of each field, which is invariant to the feature values of individual objects (Richter, Lins, & Schöner, 2021). During the grounding phase, the *expected relation fields* form peaks in dimensions centered around the reference object's feature values. The *inverse transformation fields* receive inputs from the *reference fields* and the *expected relation fields* and represent the expected spatial location/size of the target object in the same dimension representing the absolute value. They provide inputs to the *relation guidance fields*.

The **Conceptual Description** sub-network keeps WM representation of the relational structure of the base scene and activates the appropriate concepts to provide search guidance during the grounding phase (Sabinasz, Richter, & Schöner, 2023). The *object-concept field* and the *relation-concept field* jointly represent the extracted concepts and the object/relation indices. The sequence generation system (Sandamirskaya & Schöner, 2010) activates nodes standing for indices serially, so that the concepts of the first selected item are jointly represented at the location where it overlaps with the input from the object index $O1$ and the relation index $R1$ to the first relation being extracted. The *relation-target field* and the *relation-reference field* are defined over the relation index dimension and the object index dimension. The *relation-target field* receives input from the *object index field* along the object index dimension, while the *relation-reference field* receives input from the *base scene map production field* along the object index dimension. Peaks in these fields represent the argument of relations differentiating their roles. Each field mentioned above provides sub-threshold excitatory input to its counterpart *production field* (connections are not shown in the figure). During the grounding phase, the *target production field* is activated sequentially, starting from $O1$, driving other *production fields* to activate feature and relation concepts of the object selected by the index. The reference object is selected in the *relation-reference production field*. The selected features and relations are then represented in the *feature and relation concepts nodes*, ultimately guiding the spatial selection decision.

The **Decision Making** sub-network activates one of the three nodes that causes different processes to emerge based on the match/mismatch detection of features and relations. For each feature and relation dimension, there exists a *match detection* sub-sub-network that receives input from the *expected feature/relation fields* and the *attended feature/relation fields*. (Grieben et al., 2020). The *match/mismatch nodes* will be ac-

tivated depending on whether these two inputs overlap. The *no decision node* is activated if one or both inputs are missing. The *match evaluation* sub-sub-network activates one of three decision nodes based on the match detection of each dimension. The *proceed search node* is activated when the match detection is completed in all four dimensions, causing the node representing the next object index in the *target production field* to be activated, starting the search for the next object. The *reject node* is activated when there is a mismatch in at least one dimension. The grounding phase starts from the beginning again, activating $O1$ in the *target production field*. The *hypothesis testing field* keeps sustained activation of all initially selected objects. In the new grounding phase, the inhibitory input from the *hypothesis testing production field* to the *spatial selection field* causes another object to be selected as the matching object to the base scene object assigned the index $O1$ than the object that was selected beforehand. The *accept node* will be activated if sufficient feature or relation matches exist. The *match memory fields* represent match decisions of critical objects/relations for each dimension. The *cued index fields* form peaks during the index identification phase and represent the object/relation indices critical to evaluating the mapping of the cued object. The needed number of matches increases as the number of critical indices increases, as the *cued index fields* inhibit the feature/relation match nodes.

The **Mental Map** sub-network jointly represents the spatial location of objects and the assigned indices (Sabinasz et al., 2023). There exists a separate map for the base and target scenes. The *target scene map production field* forms a peak where the input from the *relation-reference production field* along the index dimension overlaps with the input from the *target scene map field*, representing the spatial location of the object selected as the reference object. The *reference space field* gets input from the *target scene map production field*. Importantly, the objects that map onto each other across scenes are assigned the same index, representing the mapping between objects. As the object index dimension binds objects in different scenes as mapping objects and the mental maps bind the object index dimension to spatial location, the spatial location of the target scene object can be identified by spatially cueing the base scene object.

The **Cognitive Control** sub-network modulates the attention given to feature and relation dimensions during the grounding phase. The *attend feature* and the *attend relation* nodes are recurrently connected to the *feature relay fields* and the *relation relay fields* respectively (Buss & Spencer, 2014). In addition, these two nodes inhibit each other mutually. During the grounding phase, the *feature relay fields* and the *relation relay fields* can form peaks only if the respective *attend nodes* are active. The *attend nodes* get external excitatory input modeling the task input. Depending on the connection strength of mutual inhibition and self-excitation, the *attend nodes* might get activated by the input from the *relay fields*.

The **Behavioral Organization** sub-network controls which fields can form peaks by providing homogeneous excitatory inputs at the right moments. Each processing phase is controlled by a set of intentions, Condition of Satisfaction (CoS), and memory nodes (Richter et al., 2021). The intention node boosts the activation of fields engaged in the intended processing phase. When the condition to terminate the process is met, the CoS node is activated, which in turn inhibits the coupled intention node. The intention nodes are chained to autonomously generate a sequence of processing phases.

## Simulation

For numerical simulation of the model, the DFT software CEDAR (Lomp, Richter, Zibner, & Schöner, 2016) was used. The condition with a featural distractor and two relations to be mapped are explained in detail, but the model can also solve tasks with one relation to map or without a featural distractor. We simulated two model variants, changing only one parameter to account for the difference between younger and older children. For the 'young' model, the mutual inhibition between the *attend nodes* is set lower than the 'old' model. This effectively causes the 'old' model to be better at suppressing the activation of irrelevant (feature) dimensions during the grounding phase. The rest of the section explains the difference in the time course generated by the young and the old model.

At the start of the simulation, both the old and the young models **describe** the base scene picture in the same way. The sustained activation in the conceptual description after the termination of the description phase is shown in Figure 4(c). The *spatial selection field* is boosted by input from the *intention describe node*, and the cued object ($O1$) is selected due to the additional excitatory input from the *cue bias*. Its features are described in the conceptual description sub-network. Next, the object to the right of the cued object ($O2$) is selected. In addition to its features, the relations between the selected and cued objects ($R1$) are described. Finally, the object to the left of the cued object ($O3$) is selected, and its features and relations to the cued object ($R2$) are described.

The *Inhibition-of-return field* is inhibited, and peaks representing the base scene objects decay. The object index $O1$ and the relation indices $R1$ and $R2$ are identified as critical indices to evaluate the mapping. Afterward, the **grounding phase of the old model** starts. Figure 4(a) shows the time course of activation changes. At $t1$, the *target production field* is activated at the location representing the object $O1$ and activates the *feature concepts nodes* standing for color 'blue' and shape 'pentagon'. The input from the *target production field* to the *relation-target production field* does not overlap anywhere with the input from the *relation-target field*, so no relation index is selected. The *attend relation node* is active due to the strong excitatory input, modeling the task instruction to focus on relation dimensions. Due to the strong inhibition from the *attend relation node*, the *attend feature node* is suppressed, and no peaks form in the *feature relay fields* despite

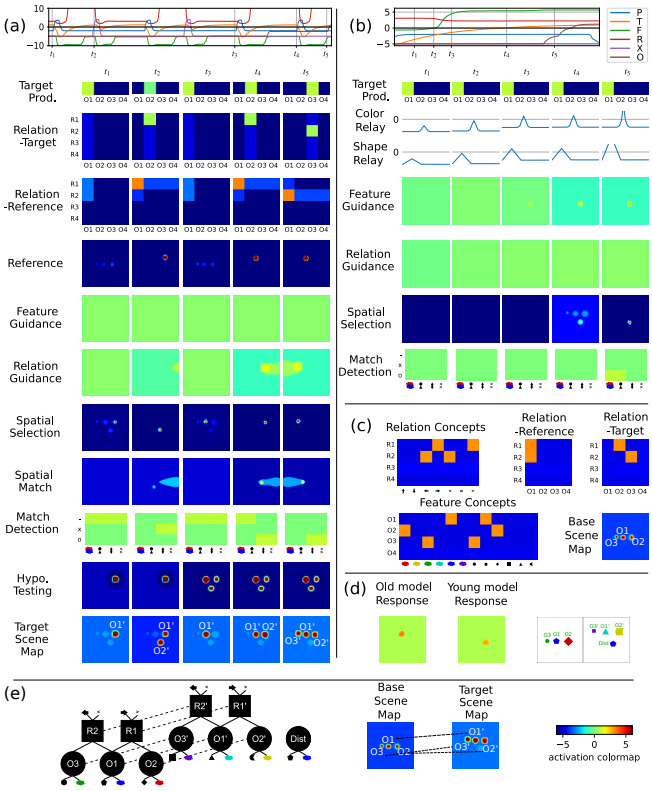


Figure 4: The activation colormap on the bottom right represents the activation value of fields. The output of 3D fields is contracted along the index dimension with labels of each object for visualization. The legend on the top right is shared for plots (a) and (b). P: *proceed search*, T:*target production delayed*, F:*attend feature*, R:*attend relation*, X:*reject*, O:*accept*. (a) The time course of activation changes in the *old* model. (b) The time course of activation changes in the *young* model. (c) The activation value after the description phase. (d) Spatial responses generated by the two models in the *target scene map production field*. (e) A graphical representation of structure mapping is on the left, and the corresponding representation based on neural dynamics in the model is on the right.

excitatory inputs from the *feature concepts nodes*. Therefore, the peak in the *spatial selection field* forms only based on saliency with no bias from the *guidance fields*. The *match detection* network decides that no match decisions can be made as no relations can be extracted yet, and there were no expected features due to the inhibition from the *attend relation nodes*. At $t2$, an object mapping to $O2$ in the base scene is being searched. The model searches for objects with a red color and a diamond shape, but there are no such objects. Likewise, the relation guidance biases the selection to the right of the first selected object and any objects bigger than the first selected object, but there are no such objects. The *spatial selection field* again selects an object based on saliency. The *match detection* network detects a mismatch in relation dimensions, and the *reject node* is activated, starting the grounding phase from the start again. At $t3$, the correct object (object mapping onto the cued object) is selected as the matching object to $O1$, as other salient objects will be inhibited from the selection due to the input from the *hypothesis testing production field* to the *spatial selection field*. At $t4$, the target index $O2$ is active. In addition to the features of $O2$, the relation $R1$ between the object $O1$ and the object $O2$ also provides guidance. The *relation guidance field* forms a peak at the location of the object that is to the right of and is bigger than the first selected object. The *match detection* network decides a match in relation dimensions. The *proceed search* node is activated, and the search for the matching object to $O3$ starts. At $t5$, an analogous process to $t4$ happens but for the target index $O3$ and the relation index $R2$. The relations between the selected and previously selected objects match the expected relations. The *accept node* is activated, and the spatial response is generated, creating a peak at the location of the object in the target scene that correctly maps onto the cued object in the base scene (left in Figure 4(d)). Two *scene maps* represent the correct structure mapping as objects with the same indices are matching objects (Figure 4(e)).

Analogous to the old model, the **the grounding phase of the young model** starts with the target index $O1$ being activated (Figure 4(b)). However, due to the weak mutual inhibition between two *attend nodes*, the *attend feature node* can get activated, ultimately making the selection decision based on feature guidance. At $t1$, the *feature relay fields* have sub-threshold bumps at the location representing feature concepts of object $O1$ due to the patterned excitatory input from the *feature concepts nodes*, but the *attend feature node* is below the detection instability, only providing weak excitatory input to the *feature relay fields*. At $t2$, the sub-threshold bumps in the *feature relay fields* and the activation level of the *attend feature nodes* have increased due to recurrent excitatory couplings between these two fields, providing weak input to each other. At $t3$, both the *attend feature node* and the *feature relay fields* go through the detection instability, recurrently exciting each other. The activation in the *feature guidance fields* starts to build up. At $t4$, the delayed excitatory input from the *target production field* to the *spatial selection field* signals a selection decision to be made, and a peak is formed in the *spatial selection field*. Due to the bias from the *feature guidance fields*, the distractor object is selected. At $t5$, the *match detection* network detects the match in two feature dimensions, and the *accept node* is activated shortly after no detection decisions have been made for the relation dimensions. The model generates a wrong response by selecting the distractor object as the object that maps onto the cued object (right in Figure 4(d)).

## Discussion

We presented a neural process model of visual analogical mapping that can build a structured representation of the visual stimuli to guide the search for matching objects, reject hypotheses based on the task requirements to focus on relational similarity, and ultimately generate a response by spa-

tially selecting a matching object to the cued object. Three key problems were addressed for this purpose.

First, the model represents the relational structure of the scene description such that the identities of relations are distinguished and the arguments of each relation are specified. In Figure 1, the relation between $O1$ and $O2$ is not confused with the relation between $O1$ and $O3$, and the relation that $O2$ is to the right of $O1$ is not confused as $O1$ being to the right of $O2$. Joint representations including the *index dimensions* enables **instantiating** objects and relations as separate entities, even when multiple entities share the same semantics.

Second, generating a spatial response was enabled by representing the established mapping between objects. Here, the *index dimension* again plays a key role in representing the structure mapping as mapped objects are assigned the same object index (*base/target scene maps* in Figure 3). The joint representation of an object's spatial location and index enables selecting a spatial location based on the cued index or vice versa. The mapping established by the model is in accordance with the structure-mapping theory as the mapping is *structurally consistent* (Gentner, 2003). Any objects in the base scene map onto at most one object in the target scene as the *inhibition-of-return field* inhibits the previously selected object from being selected again within a hypothesis. Matching relations have matching objects as their arguments, as objects are selected based on guidance from the relation to which they are bound.

Third, both the featural and the relational similarity may affect the mapping, which forms a basis for explaining why younger children are more likely to select the featural distractor as the matching object. While both feature/relation guidance guides the spatial selection, the old model could suppress the feature guidance according to the task input, while the young model failed to do so. The inhibitory connection strength between dimensional attention control nodes was set to a smaller value in the young model. This is in agreement with the hypothesis in Richland et al. (2006), stating that relatively less developed inhibitory control in younger children might explain why they are more likely to select a featural distractor. Manipulation of connection strength between dimensional attention nodes has previously been used within DFT to account for the fact that younger children are more likely to make preservative errors in the Dimensional Change Card Sort task (Buss & Spencer, 2014), which is a task hypothesized to involve inhibitory control. In general, strengthening of lateral inhibition (and local excitation) might be a general developmental change that can account for effects from other areas of cognition such as spatial WM (Schutte, Spencer, & Schöner, 2003), and effects of such development can emerge from repeated experiences (Perone & Spencer, 2013).

While our developmental account focused on changes in inhibitory control, other factors are known to influence analogical mapping. Younger children are more likely to fail at analogical mapping in which more relations have to be simultaneously considered. This might reflect limitations in their WM capacity (Halford, Wilson, & Phillips, 1998). In DFT, the WM capacity limit arises from how much activation can be sustained within fields (Perone, Simmering, & Spencer, 2011). Explanations of increased analogical mapping ability based on non-maturational factors are possible. Gentner and Rattermann (1991) claims that learning the relevant relations in the problem domain leads to increased performance which can happen on a relatively short time scale. Our model is compatible with this hypothesis, as the patterned synaptic connections between the *relation relay fields* and *relation concepts nodes* can be learned based on Hebbian rules.

Other process models of analogical mapping have been proposed. The class of models using localist representation, such as LISA (Hummel & Holyoak, 1997) and DORA (Doumas, Puebla, Martin, & Hummel, 2022), is closely aligned with our goal as they both hypothesize a sequence of cognitive processes and aim to be constrained by neural principles. In LISA and DORA, the structured representation of one scene guides the selection of matching entities in another scene. In our model, the description of one scene also guides the spatial selection of objects in another scene. This is the search strategy most used by adults. It is also used by children when solving scene analogies (Guarino, Wakefield, Morrison, & Richland, 2022). LISA and DORA represent the relational structure by instantiating a separate unit for each object-role binding. This differs from the representation in the present model, which uses the *index dimension* as an additional binding dimension (Sabinasz et al., 2023). The mapping between entities is represented in LISA/DORA by learnable connections between localist units. Our model represents the mapping by sustained activation in two mental maps. Visual attention is an important component of our model. It generates the whole process of analogical mapping from describing visual input to responding spatially. While a recent version of DORA connects to a visual preprocessor that takes visual input, that preprocessor is intended as a shortcut rather than a genuine model of visual processing (Doumas et al., 2022). In LISA and DORA, WM capacity is limited because different localist units have to share a limited amount of time to be active while being out of synchrony. These models do not seem to have an account for how items may be lost from WM. In dynamic fields, peaks may become unstable when inhibition outweighs self-excitation providing an account for how items may be lost from WM.

Morrison, Doumas, and Richland (2011) accounted for how different factors of development may influence analogical mapping ability based on LISA. While our model may be open to the influence of these different factors on analogical mapping, this has not been worked out in practice. How inferences may be based on an established analogical mapping is another interesting potential extension of our model.

## Acknowledgments

This work was supported by grant RPG-2021-350 from the Leverhulme Trust.

## References


Buss, A. T., & Spencer, J. P. (2014). The emergent executive: A dynamic field theory of the development of executive function. *Monographs of the Society for Research in Child Development*, *79*(2), vii, 1–103. doi: 10.1002/mono.12096

Doumas, L. A. A., & Hummel, J. E. (2005). Approaches to Modeling Human Mental Representations: What Works, What Doesn't, and Why. In Holyoak, Keith J. & Morrison, Robert G. (Eds.), *The Cambridge handbook of thinking and reasoning.* New York, NY, US: Cambridge University Press.

Doumas, L. A. A., Puebla, G., Martin, A. E., & Hummel, J. E. (2022). A theory of relation learning and cross-domain generalization. *Psychological Review*, *129*, 999–1041. doi: 10.1037/rev0000346

Gentner, D. (2003). Why we're so smart. In Gentner, Dedre & Goldin-Meadow, Susan (Eds.), *Language in mind: Advances in the study of language and thought.* Cambridge, MA, US: Boston Review.

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley interdisciplinary reviews: cognitive science*, *2*(3), 266–276. doi: 10.1002/wcs.105

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In Gelman, Susan A. & Byrnes, James P. (Eds.), *Perspectives on language and thought: Interrelations in development.* New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9780511983689.008

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277–300. doi: 10.1016/S0364-0213(86)80019-2

Grieben, R., Tekülve, J., Zibner, S. K. U., Lins, J., Schneegans, S., & Schöner, G. (2020). Scene memory and spatial inhibition in visual search: A neural dynamic process model and new experimental evidence. *Attention, Perception, & Psychophysics*, *82*(2), 775–798. doi: 10.3758/s13414-019-01898-y

Guarino, K. F., Wakefield, E. M., Morrison, R. G., & Richland, L. E. (2022). Why do children struggle on analogical reasoning tasks? Considering the role of problem format by measuring visual attention. *Acta Psychologica*, *224*, 103505. doi: 10.1016/j.actpsy.2022.103505

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–831. doi: 10.1017/S0140525X98001769

Hesse, M. E., Sabinasz, D., & Schöner, G. (2022). A Perceptually Grounded Neural Dynamic Architecture Establishes Analogy Between Visual Object Pairs. In *Proceedings of the 44th annual conference of the cognitive science society.* Austin, TX, USA: Cognitive Science Society.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427–466. doi: 10.1037/0033-295X.104.3.427

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lomp, O., Richter, M., Zibner, S. K. U., & Schöner, G. (2016). Developing Dynamic Field Theory Architectures for Embodied Cognitive Systems with cedar. *Frontiers in Neurorobotics*, *10*. doi: 10.3389/fnbot.2016.00014

Markman, A. B., & Gentner, D. (1993). Structural Alignment during Similarity Comparisons. *Cognitive Psychology*, *25*(4), 431–467. doi: 10.1006/cogp.1993.1011

Morrison, R. G., Doumas, L. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science*, *14*(3), 516–529. doi: 10.1111/j.1467-7687.2010.00999.x

Perone, S., Simmering, V. R., & Spencer, J. P. (2011). Stronger neural dynamics capture changes in infants' visual working memory capacity over development. *Developmental science*, *14*(6), 1379–92. doi: 10.1111/j.1467-7687.2011.01083.x

Perone, S., & Spencer, J. P. (2013). Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00648

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, *13*(4), 453–478. doi: 10.1016/S0885-2014(98)90003-X

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, *94*(3), 249–273. doi: 10.1016/j.jecp.2006.02.002

Richter, M., Lins, J., & Schöner, G. (2021). A neural dynamic model of the perceptual grounding of spatial and movement relations. *Cognitive Science*, *45*(10), e13045. doi: 10.1111/cogs.13045

Sabinasz, D., Richter, M., & Schöner, G. (2023). Neural dynamic foundations of a theory of higher cognition: The case of grounding nested phrases. *Cognitive Neurodynamics*. doi: 10.1007/s11571-023-10007-7

Sandamirskaya, Y., & Schöner, G. (2010). An embodied account of serial order: How instabilities drive sequence generation. *Neural Networks*, *23*(10), 1164–1179. doi: 10.1016/j.neunet.2010.07.012

Schöner, G., Spencer, J. P., & DFT Rearch Group. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.

Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the Dynamic Field Theory: Working Memory for Locations Becomes More Spatially Precise Over Development. *Child Development*, *74*(5), 1393–1417. doi: 10.1111/1467-8624.00614