Title
A Perceptually Grounded Neural Dynamic Architecture Establishes Analogy Between Visual Object Pairs

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Hesse, Matthis E
Sabinasz, Daniel
Schöner, Gregor

Publication Date
2022

Peer reviewed

A Perceptually Grounded Neural Dynamic Architecture Establishes Analogy Between Visual Object Pairs

Matthis E. Hesse[1,2] **(matthis.hesse@ini.rub.de)**
Daniel Sabinasz[1] **(daniel.sabinasz@ini.rub.de)**
Gregor Schöner[1] **(gregor.schoener@ini.rub.de)**
[1] Institut für Neuroinformatik, [2] Fakultät für Philosophie und Erziehungswissenschaft, Ruhr-Universität Bochum, Germany

Abstract

Detecting analogy is an important high-level cognitive skill that is involved in many aspects of human reasoning. While Structure Mapping Theory (Gentner, 1983) is a well-recognized high-level theory of analogy, it lacks a neural process implementation that links to perception and attention. Avoiding algorithmic computation on ungrounded symbols, we present a dynamic neural architecture built from interacting neural populations that establishes analogy between objects in two visually presented scenes. Consistent with SMT, it accounts for how humans find such analogies.

Keywords: analogy; dynamic field theory; neural process model; grounded cognition; embodied cognition

Introduction

Analogical reasoning is the human competence to transfer knowledge from one scene (usually a familiar situation) to another scene (a new situation), even if both settings are from different domains (Gentner & Maravilla, 2018).

A simple form of semantic analogy is expressed by the following sentence: "A is related to B as C is related to D". Here, the relation between A and B (the *target*) is explained by referring to the relation between C and D (the *base*). A famous example is the analogy between the solar system and Rutherford's model of the atom. How do we understand the analogy? We use the shared structure of the two scenes to identify the roles of entities in either of the scenes: The planets are related to the sun as the electrons to the nucleus – e.g., because the planets orbit the sun and are smaller than the sun, just as the electrons orbit the nucleus and are smaller than the nucleus. Importantly, the analogy cannot be found merely by comparing the planets with the electrons and the sun with the nucleus based on superficial similarity. Instead, it has to be identified that the planets bear to the sun the same relationships (*orbiting*, *smaller than*) as the electrons to the nucleus.

We focus on visual analogies formed on the basis of two visually presented scenes – a *base scene* and a *target scene*. In the specific task we model, each scene contains two objects that bear various relationships to each other (Figure 1). An analogy exists between the two scenes when the objects in the base scene bear the same relationships to each other as the objects in the target scene. We postulate that more than one relationship must match across the two scenes to imply an analogy (Figure 2). If an analogy exists, each object in the target scene can be mapped onto an object in the base scene so that their roles in the relationships match. The goal of the model is to provide this mapping or to conclude that no analogy exists.

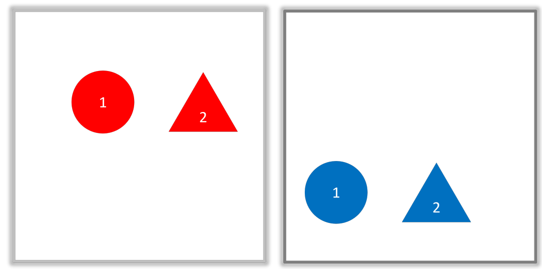


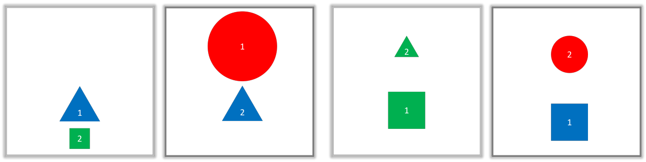
Figure 1: A typical example of a base scene (left) and a target scene (right) between which an analogical mapping can be established. Here, base scene object 1 is related to base scene object 2 as target scene object 1 is related to target scene object 2. The outer frames and the labels are for reference only and are not part of the actual visual input.

To simplify, we assume objects vary only in color, shape or size an consider only spatial relations (e.g., *left of*, *above*, ...), size relations (*smaller than*, *same size as*, *larger than*) and categorical identity relations (e.g., *same shape as*, *different shape than*). The model can be extended to more complex features and relations without altering its core.

In Figure 1, base scene object 1 bears to base scene object 2 the same relationships (*left of*, *same color as*, ...) as target scene object 1 bears to target scene object 2, thus making 1 analogous to 1 and 2 analogous to 2. The fact that base scene object 1 is also visually similar to target scene object 1 (and 2 to 2) is not relevant for the presence of an analogy, although it may aid the process of identifying the analogy.

Our goal in this paper is to propose a neural process model that may detect an analogy and establish the mapping. Avoiding algorithmic computation on ungrounded symbols, the model is based on perceptually grounded representations (Barsalou, 2008) and neural principles formalized in Dynamic Field Theory (Schöner, Spencer, & Research Group, 2016). We are guided by Structure Mapping Theory (Gentner, 1983), a widely accepted theoretical framework that describes the goals of and steps toward successful analogical inference and that is able to explain a large body of empirical data. We further back up various modeling choices

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.



(a) cross-mapping (b) no analogy

Figure 2: Two more pairs of base scene and target scene. (a) Cross-mapping: Mapping based on superficial similarity would map base scene object 1 to target scene object 2 because of the same color, shape, and size, whereas structure mapping would map 1 to 1 and 2 to 2 because base scene object 1 bears to base scene object 2 the same relationships (*above* and *larger than*) as target scene object 1 to target scene object 2. (b) No meaningful analogy exists because objects 1 and 2 share only one relationship (*above*).

by qualitative effects from the experimental literature.

Background

Structure Mapping Theory (SMT; Gentner, 1983) describes analogy as a mapping from objects in a base scene to objects in a target scene. It further suggests how analogous structure is to be found. Given a target scene and a base scene (retrieved from memory), a mapping process finds the shared systematic structure between the two scenes to establish a one-to-one mapping of all relevant entities.

Each scene includes objects which have different features and stand in relations with each other. The aim is to find the “best” map from objects of the base scene to objects of the target scene that would enable inference about the target scene from knowledge about the base scene. A central assumption is that only the relations between objects in a scene and relations between their features are relevant, and that the objective for the mapping is to preserve as many relations as possible; feature values themselves do not contribute. This process is called *structure mapping*.

How humans perform in analogy tasks differs dependent on the setup and configuration of the scenes and the task. Gentner and Toupin (1986) found that superficial similarity helps children to correctly map analogous objects. Gentner and Maravilla (2018) refer to this as the “transparency” of the mapping. Relatedly, children may be distracted into mapping objects based on superficial similarity (Loewenstein & Gentner, 2005; Richland, Morrison, & Holyoak, 2006). This effect is known as *cross-mapping* (Figure 2 (a)).

Language and concept knowledge play an important role in the discovery of analogy. Loewenstein and Gentner (2005) found that providing children with spatial language cues (such as “on top” or “below”) promotes their tendency to use relational mapping and therefore to detect analogy. Richland et al. (2006) inferred from their results that having more knowledge of relations and relational concepts may be the reason why older children make fewer errors when finding analogies. In mapping tasks conducted with children by Christie and Gentner (2014), participants performed much better when they knew terms for the presented relations and feature values. These observations motivate our assumption that feature values and relations are represented as concepts when used to detect analogies.

Methods

Dynamic Field Theory (DFT; Schöner et al., 2016) is a theoretical framework for designing neural process models. The core elements of DFT and building-blocks for neural dynamic architectures are dynamic neural fields (DNFs) that model neural populations. In these fields, a time-dependent activation $u(\vec{x},t)$ is assigned to each location $\vec{x}$ in some feature space. The activation essentially emerges from the following dynamical system:

$$\dot{u}(\vec{x},t) = -u(\vec{x},t) + h + s(\vec{x},t) + \int g(u(\vec{x}',t))k(\vec{x}-\vec{x}')\,d\vec{x}'$$

$s(\vec{x},t)$ formalizes an external input at position $\vec{x}$ and time t. g is an activation function. h is a constant negative resting level. k represents the lateral interaction of the activation dependent on the distance between field positions.

Each field generates an output $g(u(\vec{x},t))$. g is a monotonically non-decreasing function, returning values close to 0 for activation below 0 and close to 1 for activation above; it has its inflection point at 0.

The interaction kernel k takes the distance between two positions inside the field and returns the strength of interaction. Positive values result in lateral excitation and negative values result in inhibition. Usually, these interaction kernels are weighted sums of Gaussian bells, realizing local excitation, mid-range inhibition, and global inhibition. Dependent on the desired properties of the field, the parameters (amplitudes, widths, and the global inhibition constant) need to be adapted (and possibly set to 0 to not appear at all).

Relevant information in a DNF is represented via the presence of peaks. The lateral interaction together with the activation function induces instabilities: With no activation above the threshold, the system converges to its attractive and stable *sub-threshold solution* $u = h + s < 0$, no peak. When activation at certain positions in the field is above threshold (super-threshold), the system converges to its attractive and stable *super-threshold solution* $u = h + s + g(u)$ and local excitation results in a peak of super-threshold activation. This peak is *self-stabilized*: An input that led to a sub-threshold solution earlier now possibly results in a super-threshold solution, as the lateral interaction contributes to the activation. c By increasing lateral inhibition, fields can be tuned to allow only a limited number of peaks, which can serve to model limited cognitive capacities or the selection between different possible peak positions. By increasing local excitation, fields can be tuned to serve as a short-term memory, so that peaks remain even without any input. External inhibition is needed to remove peaks in such fields.

Information of any kind – such as perceptual information of a recognized object, spatial attention at a specific location, remembered object locations, values in a concept space (e.g. color, size), or relative positions – is represented by peaks in a field defined over an appropriate feature space.

Dynamic Neural Nodes hold only one value of activation. Their dynamics are similar to the dynamics of fields. Lateral interaction reduces to self-excitation. A set of neural nodes inhibiting each other resembles a field over a discrete feature space. Nodes are used to represent categories, here in the form of concepts (color, shape, size, relations). Similar to fields, there are memory nodes that may remain in their "on"-state when there is no input.

Different components can be coupled by adding some function of the output $g(u)$ of one to the input s of another to enable the construction of complex feature-rich cognitive architectures.

Model

Following a common paradigm (e.g., Loewenstein & Gentner, 2005), we assume that the base scene is presented first, followed by a presentation of the target scene in which analogical matches are to be found.

This requires keeping a representation of the base scene in short-term memory. We hypothesize that this representation is of discrete/categorical nature, i.e., that the base scene is described in terms of which feature concepts characterize the objects (e.g., *red*, *circle*, *big*, ...), and which concepts characterize the relationships the objects bear to each other (e.g., *larger than*, *left of*, ...). In the context of analogy, this choice is justified, as discussed above. For Figure 1, the base scene description would thus store the conceptual information expressible by the phrase "small red circle left of small red triangle; same size, different shape, same color".

Afterwards, a mapping is established, which is guided by the conceptual base scene description. That candidate mapping is then evaluated with respect to a goodness-of-fit criterion, which roughly corresponds to the number of matching relationships. When a mapping is accepted, a description of the analogy can be generated by removing from the base scene description all non-matching concepts. For Figure 1, that description would be expressible as "small circle left of small triangle; same size, different shape, different color".

Architecture

We combined and extended different mechanisms from the DFT framework to devise a neural process model. A very simplified overview is given in Figure 3. It can be understood as consisting of the following sub-systems:

The **Perception** system implements early processing of the visual input. Three-dimensional fields defined over two-dimensional space (corresponding to the visual array) and one feature dimension hold information about the objects' features at positions in space (Schneegans, Lins, & Spencer, 2016). We use fields for the features size, color, and shape.

The **Attentional Selection** system (Schneegans, Spencer, & Schöner, 2016) is responsible for object selection during both base scene description and search for analogical matches in the target scene: The *Spatial Attention Field* is defined over the spatial dimensions with input from the *Perception* fields, and is tuned to be selective. This models attention to one object at a time. A memory field for "inhibition-of-return" keeps track of already selected objects to avoid selecting the same object base scene object twice; a second "inhibition-of-return for target scene processing" is used for hypothesis testing when processing the target scene, to prevent that objects are processed in the same order twice. When processing the target scene, a selection represents a hypothesis: the selected object is the analog to the first (resp. second) object of the conceptually described base scene.

In the **Feature Extraction** system, attention is combined with the perception fields in a "feature/space attention field" to extract feature values of attended objects. These are then translated into a conceptual representation via nodes (Richter, Lins, & Schöner, 2021).

Relation Detection is done between two objects within one scene. Four relation detection mechanisms get input from the *Attentional Selection* end *Feature Extraction* systems and store a target and a reference location / feature value. For spatial and size relations, the relative position/value of the target as compared to the reference is determined via a steerable neural map (Schneegans & Schöner, 2012) and translated into conceptual representations of relational concepts (Lipinski, Schneegans, Sandamirskaya, Spencer, & Schöner, 2012); for color and shape there is a comparison mechanism resulting in a conceptual representation of *same* or *different*.

The conceptual **Base Scene Description** stores information about the base scene in short-term memory in the form of concept nodes (Richter et al., 2021). These nodes are memory nodes (holding their activation even in the absence of any input). For each object there is a node for every concept we account for, whose activation reflects that the respective object is described by the respective concept. These nodes are getting input from the *Feature Extraction* system. Additionally, there is one node for every relational concept we account for, whose activation reflects that the two objects stand in that relation. These nodes get input from the *Relation Detection* system. Effectively, the transient feature-based representation of objects and relationships is thereby converted into a conceptual description in short-term memory.

The way in which the target scene is processed is guided by the *Base Scene Description*. Whether the first object's analog or the second object's analog is currently searched for is controlled by a **Control and Gating Mechanism** controllable via nodes. This is a straightforward combination of nodes that passes information only when activated. Here, it is responsible for controlling (1) the *Attention Bias* and (2) the *Evaluation* (see below).

The **Attention Bias** system biases the spatial attention field towards selecting target objects that match specified features

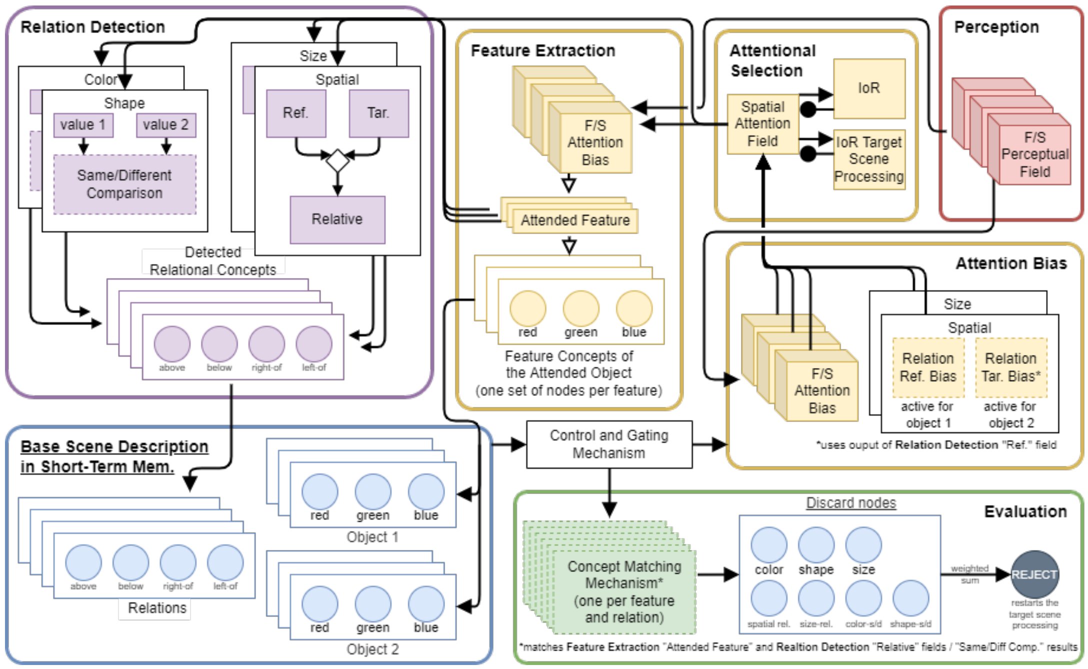


Figure 3: A simplified overview over the architecture. Not all connections are shown. The visual input goes into the perception system (top right). The outcome of analogy detection and structure mapping is represented by the nodes marked in light blue, the base scene description and the discard nodes. The **Process Organization** component of the neural architecture is not displayed; it sequentially activates and deactivates fields/nodes that may contribute and controls some of the coupling terms via gating mechanisms. This model is based on prior work, with innovation primarily in the evaluation component and the process organization, which control how hypotheses are tested, how attentional bias is induced by the base scene description and influences selection in the target scene, and how all neural processes are coordinated.

and relations. For feature-based bias, it contains one three-dimensional "feature/space attention bias" field for each feature. Each of these fields is defined over one feature dimension and two spatial dimensions, and the output of these fields provides one source of attentional selection bias. This mechanism operates in a similar way as the "scene guidance" mechanism reported in Grieben et al. (2020). For relation-based bias, we introduce a new mechanism to also promote those positions where an object could match a role (reference or target) of a specified relation. The overall attention bias has a higher value at those positions where many features and relations match, effectively promoting the selection of an analogical match that is "superficially similar" in feature values and "structurally similar" in relations. Input to the attention bias mechanisms is coming from the *Perception* system and the *Control and Gating Mechanism*.

The **Evaluation** system checks whether the selected object's features and relations (provided by the *Feature Extraction* and the *Relation Detection*) match the required concepts from the base scene description (provided by the *Control and Gating Mechanism*). For each relation and each feature concept it finds out whether the perceived feature value or relative position/value is within the concept ("match") or out of it ("mismatch"). The "mismatch"-nodes activate the so-called "discard nodes" which indicate which components of the *Base Scene Description* do not match the target scene and therefore do not belong to the common description of the two scenes. They contribute to the input of a "reject"-node. Given sufficient input, it gets activated, which represents that the selected candidate was not a good analogical match, and retriggers the selection process. To account for the higher importance of relations for a correct analogical mapping, "discard nodes" of relational concepts are weighed more strongly as input to "reject" than those of feature concepts

The **Process Organization** system controls how different subsystems of the architecture can effectively contribute. It is implemented using elementary behaviours (Richter, Sandamirskaya, & Schöner, 2012) and a serial order process organization (Sandamirskaya & Schöner, 2010). It enables to sequentially activate architecture sub-systems via an "intention"-node that is deactivated again, when the "condition of satisfaction" is met, represented by a homonymous node. In our model, there is one serial order process for the plain base scene description, and one for the target scene description.

Base Scene Processing is to be started when the model is presented with the base scene of the task. It causes an object to be selected by the *Attentional Selection* system and to be stored the *Base Scene Description* in terms of extracted feature values, and as the reference in the *Relation Detection*.

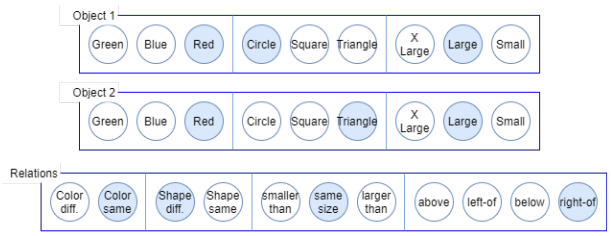


Figure 4: The *base scene description* of our example represented by memory nodes, encoding "a large red triangle right of a large red circle; having same size, having same color, having different shape".

Next, it clears the "Spatial Attention Field" and causes a new object to be selected, which is also stored in the *Base Scene Description* and as the target in the *Relation Detection*. Finally, it causes the relations of the *Relation Detection* to be stored in the *Base Scene Description*. The whole scene is described and the process terminates.

The process for **Target Scene Processing** is to be started when the input is set to the target scene and realizes the following: As a first step, the *Control and Gating Mechanism* is configured to seek for the first object. *Perception*, *Attentional Selection*, and *Feature Extraction* are reset to re-trigger the selection of objects and extraction of features. An object is selected (e.g. hypothesized to be the first base scene object's analog), it is stored as the reference in the *Relation Detection*, and the *Evaluation* is activated. If the "reject"-node stays off, the hypothesis is accepted and the process continues. The *Control and Gating Mechanism* is configured to seek for the second object, the *Attentional Selection* is cleared again, and an object is selected (e.g. hypothesized to be the second base scene object's analog) and evaluated. If the object is not rejected, the process terminates and the model succeeded to find an analogy, represented by the activated "success"-node. If the object is rejected, the whole process organization for processing the target scene can be restarted. If no more objects can be selected (since all objects have been tried), the "fail"-node will be activated, representing the failure to find an analogy between the two scenes.

Results

We tested our model on different paradigm instances and go over one example in more detail: the typical but non-trivial example already presented in Figure 1. Here, the correct mapping is to map the circle to the circle and the triangle to the triangle ($1 \rightarrow 1, 2 \rightarrow 2$). Detecting the analogy entails finding a common description. This common description does not include the color value, as it differs between the two scenes.

Processing the Base Scene

The architecture is presented with the base scene image, and the process to represent the base scene in memory is started. The attention mechanism chooses the more salient object 1 first and extracts a conceptual description that is stored in the description nodes. Then, object 2 is attended and described. The relation mechanisms extract relational concepts that are also remembered in the conceptual description nodes. (Refer to Richter et al. (2021) for a detailed visualization of a description process.) The resulting base scene description (Figure 4) includes all features and relations and is a full conceptual description of the base scene.

Processing the Target Scene

The image of the target scene is given to the architecture, and the target scene processing is started. The process is visualized in Figure 5. The first row shows the activation of the two "bias"-nodes representing that the model is seeking the analog to either the first object (in this example: a red large circle) or the second object (a red large triangle). The second row shows the activation of the success and failure nodes. When either is above threshold, the process terminates. The third row shows the activation of the "reject"-node as it sums up the mismatching features and relations. If it passes the threshold the chosen object is rejected. This plays a role in less transparent tasks (cross-mapping). The snapshots are taken at the indicated times and show: the activation of the *space attention field* representing attention to a specific location in the visual field; the activation of the *inhibition-of-return* indicating which objects were selected before; and the influence of the attention bias mechanisms on the attention selection.

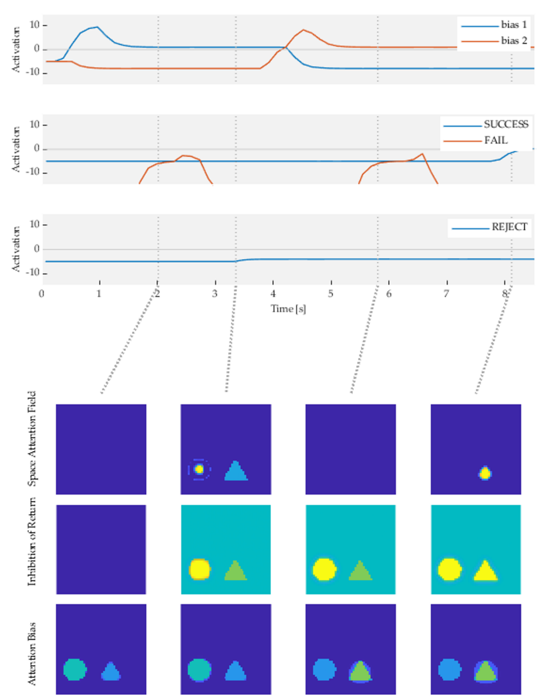


Figure 5: Visualization of some nodes and fields during the target scene processing of our example. Snapshots show the activation of fields at one time. High activation is shown by a brighter color. The last row shows the sum of all *attention bias* contributions with a different color scale.

Selecting the First Object According to the serial order, the *bias 1* is activated (representing the seeking after a first object's analog), and the perception is reset. The resulting attention bias before releasing the inhibition is shown in the first snapshot: Size ("large") promotes both objects, color ("red") promotes none of them; shape ("circle") promotes the object on the left; the spatial relation ("right-of") promotes the circle (as it can take the reference role) and the size relation ("same size") promotes both equally. The model chooses the circle, as shown in the second snapshot. It hypothesizes the analogy map $1 \rightarrow 1$. The activation of the "REJECT"-node increases, as the model detects a mismatch (color).

Selecting the Second Object To find the second object's analog, the *bias 2* is activated, and the *spatial attention field* is cleared. The *attention bias* now consists of minor selective advantage for the circle (size "large" and target role in the size relation "same size") and more selective advantage for the triangle (size "large", shape "triangle", and target role in both relations "same size" and "right of"). The last snapshot shows the successful selection of the triangle, representing the hypothesis $2 \rightarrow 2$. As not too many relations and features mismatch, the *REJECT*-node remains in its "off"-state, and the *SUCCESS*-node indicates that 0+the architecture came to a common description, i.e., it found an analogical mapping.

Response The architecture claims that it found an analogy. This analogy is given via the common description of both scenes, implicitly represented via the full *base scene description* and *discard*-nodes, indicating which part of the description does not apply to both scenes (here: color). The result is "a large triangle right of a large circle; having same size, having same color, having different shape".

We further tested our model on other examples, including those shown in Figure 2. Instance (a) is a cross-mapping example: Distracted by the superficial similarity the model first chooses the wrong object and rejects this choice only when the second object is considered. The restarted process finds the correct mapping where all relations match, although no features match. Thus, it takes the model longer and a higher-level evaluation is needed. In (b) the model does not find an analogy: it tries out both possible mappings but in each case, the non-matching features and relations are too many and the choice is rejected. In the third run of processing the target scene, no object is left and the "fail"-node is activated.

Discussion

We proposed a model that is capable of finding an analogical mapping between two visual scenes or to determine that there is none. The model is consistent with the theoretical demands of SMT and the experimental results reviewed in the Background section. It uses grounded concept representations both to create a description of the base scene and to search for analogous objects in the target scene. The model uses established neurally plausible mechanisms of short-term memory, attentional selection, process organization, visual search, relational processing, and sequence generation. Crucially, the model is a single dynamical system that generates meaningful neural representations as stable activation states that emerge from organized instabilities. Stability enables embedding the model in wider neural process accounts of grounded embodied cognition, in which sensory input may be time-varying or actively generated by gaze shifts or orienting behavior, and which may generate motor output.

This sets our model apart from accounts in which analogy is established based on ungrounded symbols (e.g., Carpenter, Just, & Shell, 1990; Lovett, Forbus, & Usher, 2010), even when such accounts are implemented using neural networks (e.g., Eliasmith & Thagard, 2001). Our goal also differs from neural network models of machine learning (e.g., Frankland, Webb, Petrov, O'Reilly, & Cohen, 2019) that learn analogy detection from examples (as evaluated by benchmarks; e.g., Webb et al., 2020), without aiming at neurally realistic models of human cognition.

Petrov (2013) describes a hybrid symbolic-connectionist model that is more closely aligned with our goals. It shares the notion that cognitive capacities emerge from interactions between component processes rather than from central processing, and provides for context sensitivity through continuous coupling to the environment. We would argue, that the neural mechanisms of that model are not fully consistent with the demands of embodiment, however.

Doumas, Puebla, Martin, and Hummel (2022) summarize extensive theoretical work based on the well-known LISA and DORA architectures that is also close in spirit and method to our effort. A groundable symbolic short-term memory of a base scene guides search for analogical objects in a target scene based on feature values. The perceptual grounding of visual features is less well embedded in neural process accounts of visual cognition, we believe, and the extent to which these models enable embodiment remains to be examined. Unlike our model, these address the learning of analogical mapping. A shared concern for both our and this work is how the hypothesized specific neural circuits scale as the number of concepts increases.

Generating a conceptual representation of the base scene makes it possible to account for the influence of language and concept knowledge as reviewed in the Background section. In the model, conceptual knowledge is critical to transfer relations from the base to the target scene. A linguistic cue to a relational concept may pre-activate the concept node.

Here we developed and demonstrated the model for a specific, very limited task. Through its mechanisms for weighing attention bias and evaluation, the model could reach a wider range of tasks and visual objects. Extending the model beyond visual structure mapping may enable linking more quantitatively to human performance (Chen, Peterson, & Griffiths, 2017). This report is is only a first exploration of how analogy may be reached in the style of DFT.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617-645. doi: 10.1146/annurev.psych.59.103006.093639

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the raven progressive matrices test. *Psychological Review*, *97*(3), 404-431. doi: 10.1037/0033-295X.97.3.404

Chen, D., Peterson, J. C., & Griffiths, T. L. (2017). Evaluating vector-space models of analogy. *arXiv preprint arXiv:1705.04416*.

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383-397. doi: 10.1111/cogs.12099

Doumas, L. A., Puebla, G., Martin, A. E., & Hummel, J. E. (2022). A theory of relation learning and cross-domain generalization. *Psychological Review*.

Eliasmith, C., & Thagard, P. (2001). Integrating structure and meaning: a distributed model of analogical mapping. *Cognitive Science*, *25*(2), 245-286. doi: 10.1016/S0364-0213(01)00036-2

Frankland, S., Webb, T. W., Petrov, A. A., O'Reilly, R. C., & Cohen, J. (2019). Extracting and utilizing abstract, structured representations for analogy. In *Proceedings of the 41st annual conference of the cognitive science society* (pp. 1766–1772). Austin, TX: Cognitive Science Society.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy*. *Cognitive Science*, *7*(2), 155-170. doi: 10.1207/s15516709cog0702_3

Gentner, D., & Maravilla, F. (2018). Analogical reasoning. In L. Ball & V. Thompson (Eds.), *International handbook of thinking and reasoning* (pp. 186–203). Routledege.

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), 277-300. doi: 10.1207/s15516709cog1003_2

Grieben, R., Tekülve, J., Zibner, S. K., 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*.

Lipinski, J., Schneegans, S., Sandamirskaya, Y., Spencer, J., & Schöner, G. (2012). A neuro-behavioral model of flexible spatial language behaviors. *Journal of Experimental Psychology: Learning, Memory and Cognition.*, *38*(6), 1490–1511. doi: 10.1037/a0022643

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315-353. doi: 10.1016/j.cogpsych.2004.09.004

Lovett, A., Forbus, K., & Usher, J. (2010). A structure-mapping model of raven's progressive matrices. In *Proceedings of the 32nd annual conference of the cognitive science society.*

Petrov, A. A. (2013). *Associative memory-based reasoning: A computational model of analogy-making in a decentralized multi-agent cognitive architecture*. LAP LAMBERT Academic Publishing.

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 for the perceptual grounding of spatial and movement relations. *Cognitive Science*, *45*(10), e13045.

Richter, M., Sandamirskaya, Y., & Schöner, G. (2012). A robotic architecture for action selection and behavioral organization inspired by human cognition. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2457-2464.

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

Schneegans, S., Lins, J., & Spencer, J. P. (2016). Integration and Selection in Multidimensional Dynamic Fields. In G. Schöner, J. P. Spencer, & the DFT Research Group (Eds.), *Dynamic thinking: A primer on dynamic field theory* (pp. 121–149). New York, NY, USA: Oxford University Press.

Schneegans, S., & Schöner, G. (2012). A neural mechanism for coordinate transformation predicts pre-saccadic remapping. *Biological cybernetics*, *106*(2), 89–109. doi: 10.1007/s00422-012-0484-8

Schneegans, S., Spencer, J. P., & Schöner, G. (2016). Integrating ' what ' and ' where ': Visual working memory for objects in a scene. In G. Schöner, J. P. Spencer, & the DFT Research Group (Eds.), *Dynamic thinking: A primer on dynamic field theory* (pp. 197–226). New York, NY, USA: Oxford University Press.

Schöner, G., Spencer, J., & Research Group, D. F. T. (2016). *Dynamic thinking: A primer on dynamic field theory*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780199300563.001.0001

Webb, T., Dulberg, Z., Frankland, S., Petrov, A., O'Reilly, R., & Cohen, J. (2020). Learning representations that support extrapolation. In *International conference on machine learning* (pp. 10136–10146).