Modelling the Sense-Making of Diagrams Using Image Schemas

Dimitra Bourou[*†], **Marco Schorlemmer**[*†] **and Enric Plaza**[*]

[*]Artificial Intelligence Research Institute, IIIA-CSIC
Bellaterra, Catalonia, Spain

[†]Dept. Ciències de la Computació, Universitat Autònoma de Barcelona
Bellaterra, Catalonia, Spain

Abstract

We model the sense-making process of diagrams as conceptual blends of the diagrams' geometric configurations with apt image schemas. We specify image schemas and geometric configurations with typed FOL theories. In addition, for the latter, we utilise some Qualitative Spatial Reasoning formalisms. Using an algebraic specification language, we can compute the conceptual blends of image schemas and geometry as category-theoretic colimits. We show through several examples how this model captures the sort of direct inferences we confer to diagrammatic representations due to our embodied cognition. We argue that this approach to sense-making might be of value for the design and application of diagrammatic and graphical visualisations, as well as for AI in general.

Keywords: diagrammatic reasoning; sense-making; image schema; conceptual blending.

Introduction

Sense-making refers to the process by which we structure our percepts into constructs that are more meaningful for us. In this work, we model the sense-making of diagrams as conceptual blends of image schemas—reflecting early embodied sensorimotor experiences (Johnson, 1987; Lakoff, 1987)—with the geometric configuration of diagrams. To the best of our knowledge, modeling the sense-making of diagrams in this manner is novel. We believe that formalising this sense-making process could be of value for fields pertaining to human-human or human-machine communication by means of graphical aids.

To illustrate our approach, take for instance the Entity-Relationship (ER) diagram of Fig. 1. Its geometric configuration comprises rectangle-shaped region boundaries *A* and *C*, a diamond-shaped region boundary *B*, and lines *ab* and *bc*, intersecting with *A* and *B*, and with *B* and *C*, respectively. Some of the possible ways one could make sense of this diagram are that the three regions form a path from *A* to *C*; or that *A* with *B*, and *B* with *C*, form pairs of entities symmetrically linked by lines. The diagram may also be thought of as consisting of three linked regions, on a path going from *A* to *C*. These conceptualizations allow the emergence of some direct diagrammatic inferences, such as an association of region *A* with *C*, due to being on the same path. Such interpretations are considered direct because they are drawn by the observers with zero inference steps (in terms of transformation steps effected on the representation). Moreover, this variety of understandings of the diagram leads to different conclusions, depending on whether the 'path' or the 'link' conceptualisation, or

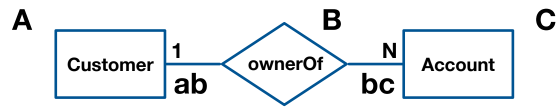

Figure 1: Geometric configuration of an ER diagram. Labels '1' and 'N' denote that one customer may have many accounts, but a single account belongs to one customer. The remaining letters label the closed curves and the lines.

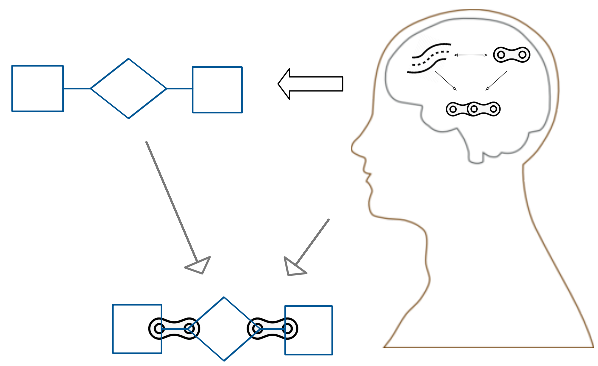

Figure 2: We distinguish the diagram geometry (left) from the diagram we make sense of (bottom). The latter arises when image schemas are blended with each other and with the geometry.

both, are at play; one imbues direction, while the other imbues symmetric association. This shows that diagrams, taken as geometric configurations, do not bring up a unique way of making sense of them.

Our proposal is to model the aforementioned sense-making process as follows: A configuration consists only of geometric entities. The ability to do inference with the diagram is not a result merely of its geometry. We claim that the sense and understanding, which make such inference possible, arise when image schemas with their rich internal structure, such as SOURCE-PATH-GOAL (PATH, in short) and LINK, are blended with each other, and subsequently with the geometry, structuring it into a meaningful diagram (Fig. 2). We model examples of such a sense-making process for diagrams from computer science and mathematics.

Background

In this section we present the theoretical background upon which our model is based.

The literature of diagrammatic reasoning has been valuable for formally studying the informational content and the efficacy of diagrams for inference. In order to reach such conclusions, a one-to-one correspondence between the geometry (the syntax) and the semantics of the diagram is typically assumed. However, as shown, a certain geometric configuration does not always evoke a unique understanding. Furthermore, the interpretation of diagrams does not simply rely on discovering a mapping, but may entail a constructive and imaginative process (May, 1999). This is in line with the process of sense-making, which we are modeling computationally in this work.

Sense-making is defined within the scope of enactive cognition as the process of an autonomous agent bringing its own original meaning upon its environment (Varela, 1991). Image schemas are fundamental for such a process, because they have the capacity to organise and structure our experience (Lakoff, 1987, p. 372). The theory of image schemas analyses such mental structures, formed early in life and constituting structural contours of repeated sensorimotor contingencies such as SUPPORT, VERTICALITY, and BALANCE. Using the latter as an example, Johnson explains that the meaning of balance emerges through embodied action and not from learning some rules (Johnson, 1987, p. 74-75). The repeated experience of different instances of balance leads to the formation of a mental structure reflecting what is invariant among them. This mental structure is a gestalt, i.e., a set of interrelated parts that make up a whole (Johnson, 1987; Lakoff, 1987). Image schemas can structure our perception and reasoning by transferring this internal structure to various domains, according to the principles of conceptual blending (Fauconnier & Turner, 2002). We model this structuring process by considering the image schemas and the geometric configuration of a diagram as constituents of a blend representing the interpreted diagram.

Conceptual blending is a process by which several mental spaces —coherent and integrated chunks of information that underlie cognition and which comprise entities, and relations or properties that characterise them—are put into correspondence with each other via cross-space relations, so as to be combined to yield a blend with novel structure (Fauconnier & Turner, 2002). In this case, the original mental spaces can be referred to as input spaces, and the resulting blend as blended space.

In this paper, we propose to model the way one makes sense of a sensory stimulus (such as a diagram), by means of cross-space correspondences between an input space reflecting this stimulus and input spaces reflecting image schemas. These correspondences enable the construction of a blended space, which integrates some substructure of each of these input spaces. Within this blended space, the emergent structure resulting from integrating the structures of the input spaces, contains novel inferences that were not possible in each input space alone.

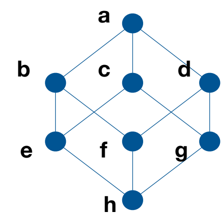

Figure 3: Geometric configuration of a Hasse diagram.

For the ER diagram of Fig. 1, the input space of the geometric configuration, comprising rectangular and diamond-shaped regions and lines, can be blended, for instance, with the input space of a PATH schema, of a LINK schema, or with the blend of both schemas. Each of these blends would yield alternative blended spaces that capture some of the different ways we can make sense of the diagram, e.g., as comprising (respectively) a directional structure, symmetrically associated entities, or a directional structure whereby each pair is symmetrically associated.

Approach

We now present, by way of several examples, our computational model of the sense-making of diagrams. We capture sense-making as the process of building the correspondences, and the resulting blend, between the geometric configurations of a given diagram, and the image schemas that provide structure to these configurations. We also show how certain direct inferences of these diagrams arise within these conceptual blends. To that end, for each example diagram we provide:

1. A formal specification of the geometry of the diagram. Qualitative Spatial Reasoning (QSR) formalisms model several aspects of spatial configurations at a level compatible with human perception (Freksa, 1991). Here, we use QSR to be able to capture the geometry of the diagram using typed first-order logic (FOL) theories.

2. Formal specifications of the structure of the relevant image schemas by means of typed FOL theories. The schemas discussed are CONTAINER, LINK, PATH, VERTICALITY, and SCALE.

3. A formalisation of suitable correspondences between image-schema structures and the geometric configuration. Given such correspondences, each conceptual blend can be modeled as a particular kind of category-theoretical colimit (Schorlemmer & Plaza, 2021).

Diagrammatic notations

We now briefly introduce the four diagrammatic notations of our case studies.

A Hasse diagram (Fig. 3) represents a partially ordered set (poset). It consists of edges and vertices, drawn as points and

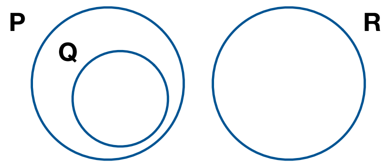

Figure 4: Geometric configuration of an Euler diagram.

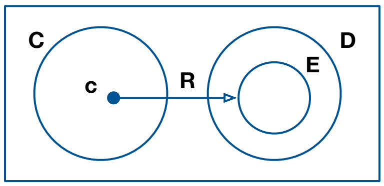

Figure 5: Geometric configuration of a Concept diagram.

lines. Each point represents one element of the poset. Assuming elements x, y and z of the poset, ordered by the '$<$' relation, then the syntactic rules are that if $x < y$, then x is shown in a lower position than y in the diagram, and that x and y are connected by a line in the diagram iff $x < y$ or $y < x$, and there is no element z such that $x < z$ and $z < y$.

ER diagrams (Fig. 1) represent entity types and relationships among them. Entity types are represented as rectangles. A relationship between two entity types is represented by a diamond, intersecting by lines with the rectangles representing the entities. In Fig. 1, the entity types involved are 'Customer' and 'Account,' and they are associated by the relationship 'owner-of.' Numbers on the lines at each side can specify how many instances of the entity of that side can participate in a relation with one instance of the entity on the opposite side (Chen, 1976).

Euler diagrams (Fig. 4) can represent inclusion and intersection relationships between sets. Their syntax consists of closed curves. Each curve represents a set, and separates the plane into two regions, whereby the interior represents the members of the set, and the exterior the non-members.

Concept diagrams (Fig. 5), an extension of Euler diagrams, can represent ontology specifications (Howse, Stapleton, Taylor, & Chapman, 2011). An external rectangle represents *Thing*, the class that all individuals are members of. Circles represent classes and dots represent individuals. The topology of both denotes their class membership, as in Euler diagrams. An arrow linking a circle (or dot) at their tail, to a circle (or dot) at their head represents a binary relation, with its domain and range. Therefore, the configuration of Fig. 5 represents that, for any individual c of class C, there is a relation R between an individual c in C and some individual in E, the latter being a subclass of D.

A formal model of sense-making

In this subsection, we discuss the formalisation of each geometric configuration and of the image schemas corresponding to it. Hasse diagrams are described in some detail by providing fragments of the actual formalisation. The other diagrams are discussed more succinctly, since the blends corresponding to them are modeled with the same principles.[1]

In order to describe the geometric configurations at hand, we draw from formal systems of the QSR literature. Regarding topology, existing formalisms enable us to characterise spatial entities as points, lines, and regions and to describe their topological relations (Egenhofer & Herring, 1991). Complex shapes can be described by identifying the points where, and the manner with which, two sides converge (e.g., acute or obtuse angle). An algorithm can automatically identify these points in digital images, and generate a qualitative description of the shape in natural language (Falomir, Gonzalez-Abril, Museros, & Ortega, 2013). The orientation of 2D objects of any shape can be formalised using eight relations (left, back-left, right, front-right, etc.) that specify the qualitative position of a primary object, with respect to a reference one (Hernández, 1991).

As for the image schemas, their structure is captured by a logical specification, based on conceptual descriptions in the literature (Johnson, 1987; Lakoff, 1987). As an example, here is a possible axiomatisation of the LINK schema in typed FOL:

$$\begin{aligned}
&\forall s \in LinkSchema : linked(anEnt(s), anotherEnt(s))\\
&\forall e_1, e_2 \in Entity : linked(e_1, e_2) \rightarrow \exists s \in LinkSchema :\\
&\quad (anEnt(s) = e_1 \wedge anotherEnt(s) = e_2) \vee\\
&\quad (anEnt(s) = e_2 \wedge anotherEnt(s) = e_1)\\
&\forall l \in Link\ \exists! s \in LinkSchema : link(s) = l\\
&\forall e \in Entity : \neg linked(e, e)\\
&\forall e_1, e_2 \in Entity : linked(e_1, e_2) \rightarrow linked(e_2, e_1)
\end{aligned}$$

Elements of type *LinkSchema* are constituted of two components of type *Entity* and one component of type *Link*, which are obtained with functions *anEnt*, *anotherEnt*, and *link*, respectively. The axioms above state that the two entities of a LINK schema are always linked; that linked entities are always part of some (not necessarily unique) LINK schema; that a link is always part of a unique LINK schema; and that the *linked* predicate is irreflexive and symmetric.

Hasse diagrams. The Hasse configuration of Fig. 3 has eight points (a to h) and twelve lines (ba, ca, etc.). Each line intersects with a pair of points. Below is a fragment of the specification of this configuration, which states the topological and orientation relations between some entities of the configuration, with respect to point a:[2]

[1] The complete formalisation of the blends modeling the sense-making of the four kinds of diagrams can be downloaded from https://drive.google.com/drive/folders/1jcQdJT0qbnAua3uXIgTEW8zV3kF_2R14?usp=sharing.

[2] Predicates such as *intersects* state the topological relations as defined in (Egenhofer & Herring, 1991), while predicates such as *right_back* state the orientation of entities as defined in (Hernández, 1991). Constants a, b, etc. are of type *Point*; constants ba, ca, etc. are of type *Line*.

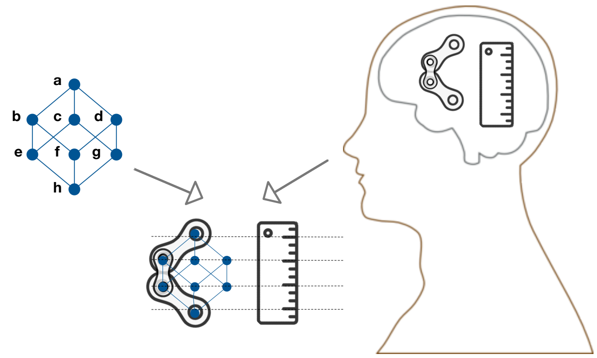

Figure 6: Conceptual blend of a Hasse diagram. The image-schematic blends CHAIN and VERTICAL-SCALE (right) are blended with the geometric configuration (left), yielding the Hasse diagram as we make sense of it (bottom).

$intersects(ba,a)$	$right\_back(a,b)$	$right\_back(a,e)$
$intersects(ca,a)$	$back(a,c)$	$back(a,h)$
$intersects(da,a)$	$left\_back(a,d)$	

The image schemas that participate in the sense-making of the Hasse configuration are LINK, PATH, VERTICALITY, and SCALE (Fig. 6). This process is modeled as the complex conceptual blend of several simpler blends (Fig. 7), discussed below.

The first blend involved is the CHAIN,[3] comprising a blend of the PATH and LINK schemas. The PATH schema relates to directed motion. It comprises a path of contiguous locations, with a source and a goal location at its endpoints. The LINK schema pertains to the notion of association, either physical or abstract. It comprises two distinct entities, linked with a link. To blend instances of these two schemas, each pair of linked entities is put into correspondence with a pair of contiguous locations in a path. Subsequently, we can compute the image-schematic CHAIN blend, whereby serially linked entities comprise a CHAIN with a source and goal as its endpoints.[4] The blend thus contains the inferences that serially linked entities on the same CHAIN are associated with each other, and that any CHAIN has two endpoints, and a direction from the source to the goal endpoint.

Ultimately, a sequence of points connected by lines in the Hasse configuration (e.g., points a, d, g, and h in Fig. 3), can be put into correspondence with a particular instance of a CHAIN by relating connected points (such as a and d, d and g, and g and h) with linked entities of CHAIN, and end points (such as a and h) with the source and the goal of CHAIN. A line intersecting with a pair of points is put into correspondence with the links of the CHAIN. More precisely, these cross-space correspondences between the mental spaces of the Hasse geometric configuration and the CHAIN blend can be expressed as pairs of a binary relation R between entities of the two spaces. For example, the sequential points a, d, g, and h and the lines da, gd and hg of Fig. 3 are related through R with components of schema instances $s_1, s_2, s_3 \in LinkSchema$ present in the CHAIN blend:

$R(anEnt(s_1),a)$	$R(anEnt(s_2),d)$	$R(anEnt(s_3),g)$
$R(anotherEnt(s_1),d)$	$R(anotherEnt(s_2),g)$	$R(anotherEnt(s_3),h)$
$R(link(s_1),da)$	$R(link(s_2),gd)$	$R(link(s_3),hg)$

These correspondences between the input spaces of CHAIN and of the geometric configuration, enable integrating them into a new blended space, as we will see later.

The Hasse configuration is also structured by a blend of the VERTICALITY and SCALE schemas (Fig. 7). VERTICALITY reflects the structure we experience from standing upright with our bodies resisting to gravity, or from perceiving upright objects like trees. Serra Borneto (1996) argues that the VERTICALITY schema comprises an axis reflecting the trajectory an object subjected to gravity would follow, or the axis of an object standing upright. In either case, a base or ground is involved as reference point. Given all the above information, we model VERTICALITY as a unique vertical axis with some marks on it, among which 'base' is the lowest one, and as a simple distinction between up and down. Lastly, the SCALE schema relates to a gradient of quantity. It comprises a structure of several grades and it has a cumulative property; if one has 15 euros, they also have 10. Consequently, we formalise SCALE as a total order on grades. Putting into correspondence the marks of VERTICALITY with the grades of SCALE schemas enables the construction of the VERTICAL-SCALE blend. This blend comprises blended levels oriented with respect to the down-up axis. The lowest of those levels corresponds to the base of VERTICALITY. An instance of VERTICAL-SCALE can be put into correspondence with the Hasse geometric configuration so that levels whereby one is ordered immediately above the other correspond to pairs of points that intersect with the same line, and one is oriented 'back' of the other.

In summary, the entire complex conceptual network of image schematic correspondences and blends, blended with the Hasse geometric configuration, yields the Hasse diagram as we make sense of it, with the following inferences: The Hasse diagram comprises several chains of linked elements, arranged at several levels of generality along a down-up vertical axis; some of these elements of different chains are on the same level of generality; there is a unique source ordered before all other elements of the diagram, and a unique goal ordered after all other elements. Mathematically, we formalise these blends as category-theoretic colimits of typed FOL theories (Schorlemmer & Plaza, 2021).

Entity-relationship diagrams. We model the sense-making of the ER diagram in an analogous way to that of the Hasse diagram, i.e., as a network of blends among the

[3] In this paper we extend the convention of typesetting the names of image schemas in small caps, to image-schematic blends as well.

[4] The resulting blended entity type is now governed by the union of the axioms of LINK and PATH in the blend.

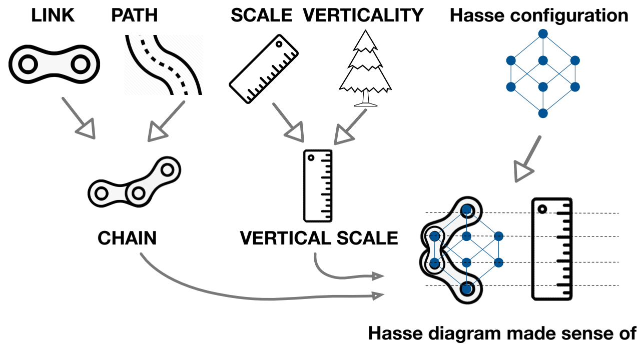

Figure 7: The network of blends modeling the sense-making of the Hasse diagram.

CONTAINER, LINK, and PATH schemas, and the ER geometric configuration. The latter has three region boundaries, two rectangle-shaped ones (A and C) and a diamond-shaped one (B), as well as two lines (ab and bc) that intersect with them in pairs (Fig. 1). The difference here is the contribution of the CONTAINER schema. CONTAINER captures the structure of entities that are hollow, and can enclose and protect other entities, ranging from a fence to a balloon. It consists of a boundary, an inside and an outside. Its structure dictates that an entity can be either in the inside or on the outside, but not both, and the transitivity of containment. Our formalisation closely follows this.

The correspondences involved between the LINK and PATH schema are the same as before, yielding a CHAIN blend. Additional correspondences are that one boundary corresponds to one entity of the CHAIN. Given these correspondences, a second blend, namely CHAIN-OF-CONTAINERS, can be constructed. Finally, CHAIN-OF-CONTAINERS is put into correspondence with the ER configuration so that one boundary corresponds to one region boundary of any shape, and that boundaries whereby one is outside of another correspond to region boundaries that are disjoint.

Consequently, an additional inference that emerges in the ER diagram, thanks to CONTAINER, is that the three region boundaries are on the outside of each other. The remaining emergent inferences are as before, i.e., region boundaries A, B, and C all become associated due to being parts of a CHAIN configuration with direction from left to right.

Euler and Concept diagrams. The modeling of these blends is done exactly as before. Here, the most prominent schema is CONTAINER. It can be inferred that some shapes are indirectly inside, or outside, some region boundaries. This is possible through the axioms in the CONTAINER specification, which are projected into the blend. In particular, in the Euler diagram of Fig. 4, region boundary Q is outside region boundary R. As for the blend capturing the sense of the Concept diagram of Fig. 5, it can be inferred that region boundary D is inside region boundary A, and point c is outside region boundaries D and E.

Related work

In this section we situate our contribution within the context of previous related work.

In diagrammatic reasoning it is often posited that the efficacy of diagrams lies in their sharing structural properties with their referents. These properties allow observers to draw direct interpretations (Stapleton, Jamnik, & Shimojima, 2017). Therefore, the more the properties of the geometry of a diagram match the properties of its semantics, the more efficacious this diagram would be to represent this semantics. Here, we expanded in this direction by modeling the origin of these properties as the blending of image schemas with the geometry of a diagram.

A few research groups have worked on formalising image schemas and the relations among them. Rodríguez and Egenhofer (2000) provide a relational algebra inspired by the CONTAINER and SURFACE schemas, used to model, and reason about, spatial relations of objects in an indoor scene. Image schemas have also been used to model planning and actions of agents St Amant et al. (2006). Some image schemas were recursively defined as compositions of other schemas. In both these works, the formalisations are inspired by image schemas, rather than faithful representations of their descriptions in the literature. Embodied Construction Grammar formalises (Bergen & Chang, 2005) and implements (Bryant, 2008) language understanding by mapping components of specific schemas (image schemas, and others) to phonemes. This work is analogous to our own, except the stimulus made sense of is a diagram and not a spoken sentence.

Kuhn (2007) formalised image schemas as ontology relations using functional programming in a relatively abstract manner. In a recent, comprehensive work, Hedblom (2020) modeled image schemas as families of interrelated theories, with each schema comprising a combination of primitive components. QSR formalisms that capture the spatiotemporal content of schemas were used. In the present approach, we chose not to use such formalisms to capture the internal structure of image schemas.

There have also been several efforts to provide mathematical models that formalise a blending process for given input spaces. Related to our approach, Goguen (2006) applied algebraic specifications and their category-theoretic operations for modeling the cognitive understanding of space and time when solving a riddle. Building on this work, Schorlemmer, Confalonieri, and Plaza (2016) modeled the creative problem-solving process of tackling the same riddle by way of a category-theoretic characterisation of blending, based on typed FOL specifications of image schemas. In a similar vein, Hedblom (2020) also implemented a small example of blending linguistic concepts using image schemas. As with the work by Schorlemmer et al., image schemas are used to establish shared structure between two input spaces. Image schemas have also been used to interpret an icon by blending a description of the schema with a QSR description of the icon (Falomir & Plaza, 2019). This approach is a conceptual equivalent of the current computational model, referring to the sense-making of icons, instead of that of diagrammatic configurations.

Discussion

In this paper we have presented a formal framework of the sense-making of diagrams, modeling the way observers structure diagrams by unconsciously projecting preexisting mental structures — i.e., image schemas — on the geometry of a diagram, giving rise to direct inferences.

We have already described some interpretations that observers can make directly when encountering specific types of geometric configurations, and which are not fully determined by that configuration itself. Here, these interpretations are made precise and formal. In the case of the Euler diagram (Fig. 4), the inference of region boundaries R and Q being disjoint is directly observable from the configuration, because of its inherent structural properties (Stapleton et al., 2017). Here, we model such direct inferences as arising in the blend. Likewise, in the Hasse diagram (Fig. 6), the transitive ordering of points in terms of their degree of the VERTICAL-SCALE schema, the inference that the source of an instance of CHAIN is ordered before all others (minimal element), and the goal is after all others (maximal element), as well as the existence of distinct instances of CHAIN (maximal chains), are all direct inferences made possible by the blending of image schemas with the geometric configuration.

Therefore, some geometric configurations are more efficacious for representing a given semantics, than others, due to having more similar properties with this semantics (Stapleton et al., 2017). For example, set membership is represented more efficaciously with enclosure in closed curves (Fig. 4) than with points of a line (Mineshima, Sato, Takemura, & Okada, 2014), because curve enclosure is transitive and asymmetric (geometrical property), as is set inclusion (semantic property). Taking one step back, our conjecture is that this correspondence between geometric and semantic properties, is explained by the fact that some geometric configurations can be integrated with image schemas that lead to inferences compatible with the intended semantics. Such examples are: the CONTAINER schema, in the case of Euler diagrams and the semantics of set inclusion, and the SCALE and VERTICALITY schemas, in the case of Hasse diagrams and the semantics of posets. The above will be further explored formally in future work.

Conclusions and Future Work

In this paper we have modeled the sense-making of diagrams, and the inferences carried out with them, as conceptual blends of an observer's embodied cognitive structures —crystallized as image schemas— with the geometry of a diagram. The novelty of our framework lies in the fact that it is not merely conceptual but also written in a formal, computer-processable language. We contribute to the literature with a reusable set of formalized image schemas. Importantly, our framework is general enough to be applicable to different types of stimuli. Therefore, although in this paper we have modeled a blending process to represent the sense-making of geometric configurations of diagrams, the stimuli made sense of could in principle be any structure that is expressible in typed FOL. Furthermore, the entire framework could eventually be generalised in a representation-independent manner as described in (Schorlemmer & Plaza, 2021).

One limitation of the current work is the lack of a quantitative metric of evaluation of the outcome of the sense-making process. However, the view of the theories of conceptual blending, image schemas and enactivism is precisely that sense-making has no ground truth, as sense arises subjectively for each organism through its experience with relation to a body and an environment. In the framework of conceptual blending, correspondences are not prescriptive and there is a variety of different ways to blend input spaces and obtain new meaning. Nonetheless, even in this subjective and relational view, a good blend would be one that is well integrated (Fauconnier & Turner, 2002, ch. 16) and, in our approach, one that leads to emergent structure containing valid inferences. More concretely, in the current work, as in most of the literature of diagrammatic reasoning, we prescribed the cross-space correspondences, and thus the blends, so that they give rise to inferences that are consistent with the intended semantics of a diagrammatic notation.

The issue of generating and evaluating alternative blends for a given configuration, including those that model erroneous interpretations, i.e., inconsistent ones with the intended semantics, will be explored in the future. To that end, we intend to model a wider range of image schemas, obtaining a reusable, comprehensive library. In another direction, we would like to extend our research to formally explore why some diagrams are interpreted faster or more accurately than others. This information is of value for human-computer interaction because it could provide guidelines for the design of efficacious diagrammatic and graphical visualizations.

Although our direct contributions pertain to diagrams, our formal framework could be useful for other areas in AI. Image schemas reflect invariants across states of affairs, enabling their abstraction into a specific concept. For example, cups of various shapes and materials, perceived through vision or touch, are a CONTAINER. It has indeed been proposed that the generalisation capabilities of AI systems regarding knowledge about the physical world, which some animals possess, could be improved through the acquisition of abstract concepts like gravity, container, and motion along a path (Shanahan, Crosby, Beyret, & Cheke, 2020; Thosar et al., 2020). Interestingly, such concepts correspond directly to image schemas, and future developments of our blending framework could model the way agents use these concepts to make sense of various states of affairs they perceive.

Acknowledgments

The present research was supported by CORPORIS (PID2019-109677RB-I00) funded by Spain's *Agencia Estatal de Investigación*; by DIVERSIS (201750E064) and DIAGRAFIS (202050E243) funded by CSIC; and by the *Ajuts a grups de recerca consolidats* (2017 SGR 172) funded by *AGAUR / Generalitat de Catalunya*.

References

Bergen, B., & Chang, N. (2005). Embodied construction grammar in simulation-based language understanding. In *Construction grammars: Cognitive grounding and theoretical extensions* (Vol. 3, pp. 147–190). John Benjamins.

Bryant, J. E. (2008). *Best-fit constructional analysis* (PhD dissertation). EECS Department, University of California, Berkeley, Magdeburg, Germany. (http://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-100.html)

Chen, P. P.-S. (1976). The entity-relationship model—toward a unified view of data. *ACM Trans Database Syst*, *1*(1), 9–36.

Egenhofer, M. J., & Herring, J. R. (1991). *Categorizing binary topological relations between regions, lines, and points in geographic databases* (Tech. Rep.). Department of Surveying Engineering, University of Maine.

Falomir, Z., Gonzalez-Abril, L., Museros, L., & Ortega, J. A. (2013). Measures of similarity between objects based on qualitative shape descriptions. *Spat Cogn Comput*, *13*(3), 181–218.

Falomir, Z., & Plaza, E. (2019). Towards a model of creative understanding: Deconstructing and recreating conceptual blends using image schemas and qualitative spatial descriptors. *Ann Math Artif Intell*, *88*, 457–477.

Fauconnier, G., & Turner, M. (2002). *The way we think*. Basic Books.

Freksa, C. (1991). Qualitative spatial reasoning. In *Cognitive and linguistic aspects of geographic space* (pp. 361–372). Springer.

Goguen, J. (2006). Mathematical models of cognitive space and time. In *Reasoning and cognition* (pp. 125–128). Keio University Press.

Hedblom, M. M. (2020). *Image schemas and concept invention: cognitive, logical, and linguistic investigations*. Springer.

Hernández, D. (1991). Relative representation of spatial knowledge: The 2-D case. In *Cognitive and linguistic aspects of geographic space* (pp. 373–385). Springer.

Howse, J., Stapleton, G., Taylor, K., & Chapman, P. (2011). Visualizing ontologies: A case study. In *The semantic web – ISWC 2011* (pp. 257–272). Springer.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press.

Kuhn, W. (2007). An image-schematic account of spatial categories. In *Spatial information theory, 8th international conference, COSIT 2007.* Springer.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. University of Chicago Press.

May, M. (1999). Diagrammatic reasoning and levels of schematization. In *Iconicity. a fundamental problem in semiotics* (pp. 175–194). NSU Press, Copenhagen.

Mineshima, K., Sato, Y., Takemura, R., & Okada, M. (2014). Towards explaining the cognitive efficacy of euler diagrams in syllogistic reasoning: A relational perspective. *J Vis Lang Comput*, *25*(3), 156–169.

Rodríguez, M. A., & Egenhofer, M. J. (2000). A comparison of inferences about containers and surfaces in small-scale and large-scale spaces. *J Vis Lang Comput*, *11*(6), 639–662.

Schorlemmer, M., Confalonieri, R., & Plaza, E. (2016). The Yoneda path to the Buddhist monk blend. In *Proc. of the Joint Ontology Workshops 2016.* CEUR-WS.org.

Schorlemmer, M., & Plaza, E. (2021). A uniform model of computational conceptual blending. *Cogn Syst Res*, *65*, 118–137.

Serra Borneto, C. (1996). Liegen and stehen in German: A study in horizontality and verticality. In *Cognitive linguistics in the redwoods* (pp. 459–506). Mouton de Gruyter.

Shanahan, M., Crosby, M., Beyret, B., & Cheke, L. (2020). Artificial intelligence and the common sense of animals. *Trends Cogn Sci*, *24*(11), 862–872.

St Amant, R., Morrison, C. T., Chang, Y.-H., Mu, W., Cohen, P. R., & Beal, C. (2006). *An image schema language* (Tech. Rep.). North Carolina State University at Raleigh, Dept. of Computer Science.

Stapleton, G., Jamnik, M., & Shimojima, A. (2017). What makes an effective representation of information: a formal account of observational advantages. *J Logic Lang Inf*, *26*(2), 143–177.

Thosar, M., Mueller, C. A., Jaeger, G., Pfingsthorn, M., Beetz, M., Zug, S., & Mossakowski, T. (2020). Substitute selection for a missing tool using robot-centric conceptual knowledge of objects. In *Proc. 35th Annual ACM Symposium on Applied Computing* (pp. 972–979).

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In *Organism and the origins of self* (pp. 79–107). Springer.