

Image Schema Combinations and Complex Events

Maria M. Hedblom · Oliver Kutz · Rafael Peñaloza · Giancarlo Guizzardi

Received: date / Accepted: date

Abstract Formal knowledge representation struggles to represent the dynamic changes within complex events in a cognitively plausible way. Image schemas, on the other hand, are spatiotemporal relationships used in cognitive science as building blocks to conceptualise objects and events on a high level of abstraction. In this paper, we explore this modelling gap by looking at how image schemas can capture the skeletal information of events and describe segmentation cuts essential for conceptualising dynamic changes. The main contribution of the paper is the introduction of a more systematic approach for the combination of image schemas with one another in order to capture the conceptual representation of complex concepts and events. To reach this goal we use the image schema logic ISL, and, based on foundational research in cognitive linguistics and developmental psychology, we motivate three different methods for the formal combination of image schemas: merge, collection, and structured combination. These methods are then used for formal event segmentation where the changes in image-schematic state generate the points of separation into individual scenes. The paper concludes with a demonstration of our methodology and an ontological analysis of the classic commonsense reasoning problem of ‘cracking an egg.’

Keywords image schemas · knowledge representation · commonsense reasoning · event structure · ontology design patterns · egg cracking problem

1 Introduction

Formally capturing the nature of complex concepts and events, and the dynamic transformations they bring about in the world, is a difficult problem. In comparison, what formal knowledge representation struggles with, humans perform without much thought or effort. Based on experiences, humans have an understanding of concepts and events (simple and complex), and can reason about outcomes, make predictions, reason backwards from an observation, and adapt their conceptualisation to changes even in unfamiliar scenarios. If there is a mismatch between a conceptualisation and an observed situation, humans can easily modify conceptualisations and re-represent the observed situation. This flexibility in mentally representing and updating information is not as straightforward for formal knowledge representation, aimed at automated reasoning. Previously, representations of the cognitive perception of real world scenes were sometimes based on formal frameworks used in naïve physics [29], such as situation calculus or causal logic. In addition, classic commonsense reasoning problems such as *cracking an egg* [45, 51] were then often described with long and complex axiomatisations that offer little in terms of cognitive adequacy or conceptual clarity. More importantly, they do not match the level of abstraction on which humans seem to reason. While such methods were quite influential within knowledge representation and computational logic, research in cognitive science has recently gained new insights and more embodied theories

This paper is an extended and improved version of [34].

M. Hedblom · O. Kutz · G. Guizzardi
Conceptual and Cognitive Modelling Research Group CORE,
KRDB Research Centre for Knowledge and Data, Faculty of
Computer Science, Free University of Bozen-Bolzano, Italy
E-mail: {mhedblom,okutz,gguizzardi}@unibz.it

R. Peñaloza
Department of Computer Science, Systems and Communica-
tions, University of Milano-Bicocca, Italy
E-mail: rafael.penalaza@unimib.it

of cognition have found their computational matches in statistics and machine learning techniques. One suggestion on how repeated human experience is cognitively structured is through generalised, mental structures. One example of those structures are *image schemas* [35, 44]. Image schemas are learned from early sensorimotor experiences, and can be found in natural language and in analogical reasoning. They are studied in cognitive linguistics (e.g. [28]), developmental psychology (e.g. [47]) and formal knowledge representation (e.g. [33]). Image schemas are often described as spatiotemporal relationships, such as CONTAINMENT¹ and SOURCE_PATH_GOAL (SPG). A concept like ‘journey’ can be conceptualised with SPG and an object like ‘cup’ with the affordance for CONTAINMENT. We argue that, by using the formal representation of such conceptual primitives in different combinations, it is possible to approach a more cognitively plausible representation of events. Initially, this formal representation needs to be bootstrapped for the most simple image schemas, for which we employ the tailor-made spatiotemporal logic for image schemas ISL, introduced in [32]. Formalisations of more complex image schemas are derived from those for simpler ones, and complex events are described as a temporal sequence of scenes carrying significantly distinct image-schematic information.

The approach as just described requires handcrafted formalisations and analysis of the event structure, and therefore does not scale well to fit applications in, e.g., cognitive robotics. However, it is possible to augment the handcrafted logical representation of image schemas with machine learning approaches detecting the satisfaction of image schematic states (see e.g. [27] for early work in this direction). Such a hybrid approach is therefore still based on the same fundamental principles of cognitively inspired modelling of events using image schemas, whilst avoiding both, handcrafted modelling of temporal event structure as well as logical modelling of causation and physics (instead relying on simulations). However, an additional problem needs to be tackled. For more complex and dynamic concepts one image schema alone usually cannot fully capture the image-schematic skeleton underlying a conceptualisation. Instead, the image schemas need to be grouped and combined with one another. Image schema combinations, sometimes called *profiles*, are commonly mentioned in the literature (see e.g. [55]), yet to our knowledge there exists no systematic method for describing these combinations. In order to contribute to this research agenda, this paper addresses the problem of image schema combinations and illustrates how their

formal representation can be used as *modelling patterns* (in the sense of the Foundational Ontology Patterns introduced in [16]) for the representation of dynamic concepts and events.

2 The Foundations of Meaning

Conceptual meaning has been suggested to be associated with uses and purposes of objects and events, rather than with their perceivable attributes and visual patterns [48, 61]. For instance, while a *cup* might be visually identified by the spatially occupied combination of a hollow cylinder with a handle, as defined through theories such as recognition-by-parts [7], it is only the affordance to contain e.g. liquid that makes it in fact a cup.

Unlike for objects, there are no ‘borders’ in the passing of time. One event often floats seamlessly into another without pauses, beginnings or ends. Despite this, events are also often distinguished by their spatial dimension. The human mind also has an ability to take dynamic perceptions and, based on certain cognitive principles grounded in spatiotemporality, identify when a new *event* takes place [40, 67]. This ability emerges already at an early stage as children learn to distinguish between different events and to make ‘conceptual cuts’ in the stream of perception (e.g. [2]). These ‘event pieces,’ which may be temporal, spatial, or material, can, in different combinations, represent increasingly complex and large-scale situations.

For instance, an event like *going to the library* can be described as ‘a person moving towards a library-building’ together with an understanding of the core *participants* therein (such as Person, Library, Road). At the same time, we associate a library and going there, to a full range of additional conceptual information such as ‘lending and returning,’ ‘book collection,’ ‘knowledge,’ ‘public place,’ etc.; namely, information that in itself is not perceptual but based on particular experience through the affordances that these particular concepts realise.

Research in cognitive linguistics also demonstrates these tendencies: there exists a range of different theories trying to explain how information is broken into smaller conceptual structures. Additionally to image schemas, semantic primes [66], and conceptual primitives [63] have been introduced as possible such frameworks. Such approaches typically do not claim a monopoly on the right choice of particular conceptual primitive, but focus on some particular explanatory goals. Therefore, our bias to the realm of image schemas is not intended to be exclusive but to be seen as a starting step in our study.

¹ Following convention, image schemas are written in uppercase letters.

Image schemas represent abstract generalisations of events usually learned from sensorimotor processes [35, 44]. They correspond to conceptual *gestalts*, meaning that each part is essential to capture the image schema,² and are commonly described as capturing sensorimotor patterns of relationships and their transformations. An important aspect is that image schemas exist in both static forms (e.g. LINK, CONTAINMENT and CENTER_PERIPHERY) and in dynamic, temporally-dependent forms (e.g. LINKED_PATH, Going_IN and REVOLVING_MOVEMENT) [9]. For simplicity and in terms of priority, many formal studies of image schemas have focused on capturing the static aspects of image schemas (e.g. [5]). However, in order to represent events and more dynamic concepts, also the temporal and transformational dimension of the image schemas require attention. Some work has been done to model the dynamic aspects of image schemas but they are often limited to a particular schema or situation that cannot be easily generalised (e.g. [21, 31]).

While image schemas such as SCALING or CYCLE implicitly contain a temporally-dependent transformation, most often more than a single image schema is required when modelling complex concepts and events. In relation to image-schematic structures, Dodge and Lakoff [14] argue that (linguistic) “*complexity and diversity can be explained in terms of combinations of simple universal primitives.*” The principle that image schemas can be combined with one another is a fundamental aspect of how they construct meaning both in natural language and in the conceptualisation of objects and events.

For this purpose, image schemas have been suggested to be gathered into ‘profiles’ which represent the full spatiotemporal skeleton for the conceptualisation of a particular concept [55]. For instance, [22] provides a plethora of image schema profiles for the word *stand* based on different linguistic contexts. Describing the image schema profile of the event *going to the supermarket*, one can use a collection of the following image schemas: SPG—as I am going to the supermarket; CONTAINMENT—as myself and the groceries are inside the building, PART_WHOLE and COLLECTION—as there are plenty of pieces in the supermarket and I collect them, TRANSFER—as I am obtaining objects from the supermarket and ‘transfer’ them to my own ‘person,’ etc. We will see this basic idea further analysed and at work below.

3 Formally Representing Image Schemas Using ISL: The Image Schema Logic

Image schemas are abstract patterns that become detectable only due to their prevalence in natural language and cognition in general. Therefore, much like with all spatiotemporal formalisation problems, it is not trivial to formally represent them in a satisfactory way [3, 20]. The landscape of logical formalisms, including spatiotemporal logics, is currently unified by the research on universal logic [23, 42], which aims to give abstract and general definitions for the notion of ‘logic’ [54] and ‘logical translation’ [53], and to produce logic-agnostic meta-results and semantic foundations for meta-languages such as DOL [52].

One problem for formalising image schemas is that the cognitive-driven investigations of how humans perceive and experience time cannot easily be mapped to existing temporal logic approaches [8, 13, 56]. These limitations to the use of off-the-shelf calculi also extend to the spatial domain. A well known formalism, which has been extensively used for the representation and handling of qualitative spatial knowledge is the Region Connection Calculus (RCC) [10]. Unfortunately, cognitive studies have supported the claim that humans do not typically make, or accept, some of the distinctions inherent to the RCC calculus [36]. Despite this potential cognitive mismatch, some research on image schema formalisation still uses RCC (see for instance [5, 21]) since it does provide a direct and easy to understand formal representation of space and associated notions such as ‘overlap’ and ‘contact.’

3.1 ISL: The Image Schema Logic

While image schemas are often discussed without an immediate formal correspondence, there exists a number of attempts to capture them formally (e.g. [5, 19, 39]). The formal language ISL [32]³ is intended to capture the basic spatiotemporal interactions which are relevant for image schemas. Briefly, ISL is an expressive multi-modal logic building on RCC [58], Ligozat’s Cardinal Directions (CD) [46], Qualitative Trajectory Calculus (QTC) [65], with 3D Euclidean space assumed for the spatial domain, and Linear Temporal Logic over the reals (RTL). The work on formalising individual image schemas and their dynamic transformations in ISL was initiated, for instance, in [31] and expanded to include agency in [43] through the addition of see-to-it-that (STIT) logic [4].

² For instance, consider a container without an inside.

³ ISL was further developed under the name ISL^{FOL} by the addition of a First-Order concept language in [30].

At its core, ISL follows a popular temporalisation strategy (studied in further detail in [18]), where temporal structures are the primary model-theoretic objects (e.g., a linear order to represent the passage of time), but at each moment of time we allow complex propositions that employ a secondary semantics. The atoms in ISL are then topological assertions about regions in space using RCC8, the relative movement of objects w.r.t. each other using QTC, and relative orientation, using CD. The purpose of quantification is to separate different sortal objects, while otherwise the syntax of the language follows a standard multi-modal logic paradigm.

We briefly sketch the sublogics that build ISL and how they are combined. We refer the reader to [30, 32, 43] for more detailed accounts of the theoretical aspects of this language and the sublogics that compose it.

The spatial dimension—topology of regions. Following, amongst others [5, 21], RCC is used to represent basic topological spatial relationships for image schemas. We in particular use the RCC8 relations [58] since a mere mereological description would not suffice for modelling image schemas. Indeed, it is important to distinguish, for example, whether two objects touch each other (EC) or not (DC).

The spatial dimension—cardinal directions. In general, directions may be absolute or relative. Usually, left and right are considered relative directions [62], which however are conceptually and computationally much more complicated than (absolute) cardinal directions [46] like North or West. Basic ISL assumes a naïve egocentric view (that is, with a fixed observer), from which directions like left/right, front/behind and above/below can be recognised as cardinal. This leads to six binary predicates on objects: *Left*, *Right*, *FrontOf*, *Behind*, *Above* and *Below*. Note that these relations are unions of base relations in a three-dimensional cardinal direction calculus as in [46], and the latter can be recovered from these relations by taking suitable intersections and complements.

The movement dimension. To take the dynamic aspects of image schemas into account, the Qualitative Trajectory Calculus (QTC) [65] is used to represent object relationships in terms of movement. This results in nine different relations. In its variant QTC_{B1D} , the trajectories of objects are described in relation to one another. We simplify the calculus by considering only the following three possibilities:

- (i) if object O_1 moves *towards* O_2 's position, this is represented as $O_1 \rightsquigarrow O_2$;

- (ii) if O_1 moves *away from* O_2 's position, this is represented as $O_1 \leftarrow O_2$; and
- (iii) O_1 being *at rest* with respect to O_2 's position is expressed as $O_1 \circ O_2$.

This approach for writing the relative movement of two objects is intuitive and expressive enough to justify its use as a representation language. With QTC, we can speak about relative movement for a given time point. What is missing is the ability to speak about temporal changes.

The temporal dimension. We use the simple linear temporal logic RTL over the reals [38, 50, 59] with future and past operators. The syntax of this logic is defined by the grammar

$$\varphi ::= p \mid \top \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \mathbf{U} \psi \mid \varphi \mathbf{S} \psi$$

where $\varphi \mathbf{U} \psi$ reads as “ φ holds, until ψ ” and $\varphi \mathbf{S} \psi$ reads as “ φ holds, since ψ .”⁴ As it is standard in temporal logic, we can define additional temporal operators based on these two; for example, operators:

- $\mathbf{F}\varphi$ (at some time in the future, φ) is defined by $\top \mathbf{U} \varphi$,
- $\mathbf{P}\varphi$ (at some time in the past, φ) is defined as $\top \mathbf{S} \varphi$,
- $\mathbf{G}\varphi$ (at all times in the future, φ) is defined as $\neg\mathbf{F}\neg\varphi$,
- $\mathbf{H}\varphi$ (at all times in the past, φ) is defined as $\neg\mathbf{P}\neg\varphi$.

ISL is constructed by combining all these languages in a controlled manner as described next.

3.2 Syntax and Semantics of ISL

The syntax of ISL is defined over the combined languages of RCC8, QTC_{B1D} , cardinal direction (CD), first-order logic and linear temporal logic over the reals (RTL), with the 3D Euclidean space assumed as the interpretation for the spatial domain. Note that we need to interpret the temporal constructors over real-time in order to handle QTC relations, whose semantics implicitly assume continuous time. Modifying components of ISL therefore requires a careful control of the global semantics.

Formally, sentences of ISL are first-order RTL temporal formulas constructed over (ground) atomic formulas taken from the union of RCC8 statements, 3D cardinal directions, and QTC_{B1D} , which we briefly introduced before, together with a standard first-order

⁴ As shown in [50], past time operators can be added to a logic with only future modalities without causing trouble regarding decidability and complexity.

application of predicates. We sketch the ISL logic as originally presented in [30] (slightly different from the presentation in [32]), assuming a basic acquaintance of the semantics of the component logics, and focusing on the semantics for the integrated logic.

ISL considers three sorts of objects, each of them interpreted as certain (further constrained) subsets of \mathbb{R}^3 . These sorts are *objects*, *regions*, and *paths*. Intuitively, objects occupy arbitrary subsets of \mathbb{R}^3 , and they denote and occupy different regions at different times. Rigid and non-rigid regions over time can be introduced, but here we only consider quantification over objects that denote rigid regions⁵ in order to stay in a first-order quantificational paradigm. More precisely, the objects that we quantify over can be seen as abstract objects, but formal models for the ISL language include an extension function that associates with any such object the region it occupies in \mathbb{R}^3 , an approach which follows the semantic paradigm of counterpart theory [37] and \mathcal{E} -connections [41].

Finally, a path is interpreted as a continuous function from the unit interval $[0, 1]$ into \mathbb{R}^3 , allowing the definition of the source and the goal of a movement along a path as the values of 0 and 1, respectively. In this version of ISL, these values are 0-dimensional 1-point subsets of \mathbb{R}^3 . Extensions of extended objects may be normalised to denote regular closed subsets of the topology of \mathbb{R}^3 in accordance with the typical usage of RCC8.⁶

For a fixed set X of object, region, and path variables and each sort s we define the set of terms $T_s(X)$ of sort s . For example, if t is a term of type ‘path,’ then $source(t)$ is of type ‘region,’ etc. (see [30] for a full definition). Given this, the set of atomic formulas are defined as:

- $t = u$ for $t, u \in T_s(X)$,
- $p(t_1, \dots, t_n)$ for $p : w \in P_r \cup P_f$ and $t_i \in T_{s_i}(X)$ for $i = 1, \dots, n$,
- $DC(t, u)$, $EC(t, u)$, $OV(t, u)$, $EQ(t, u)$, $TPP(t, u)$, $TPPi(t, u)$, $NTPP(t, u)$, $NTPPi(t, u)$,
for terms $t, u \in T_{Region}(X) \cup T_{Path}(X)$,
- $Left(t, u)$, $Right(t, u)$, $FrontOf(t, u)$, $Behind(t, u)$,
 $Above(t, u)$, $Below(t, u)$,
for terms $t, u \in T_{Region}(X) \cup T_{Path}(X)$,
- $t \rightsquigarrow u$, $t \leftarrow u$, $t \circ u$, for terms $t \in T_{Object}$, and $u \in T_{Region}(X)$.

⁵ Rigid elements are those that do not change their extension through time.

⁶ Usually, RCC8 is restricted to regular closed sets. However, for the current purposes it should cover images of paths and source and target points of paths as well, and these are not regular closed.

Finally, ISL formulas are first-order RTL formulas built over these atomic formulas in the usual way. Moreover, satisfaction of complex formulas is inherited from RTL: φ holds in M , denoted $M \models \varphi$, if for all time points $t \in \mathbb{R}$ and all valuations $\nu : X \rightarrow M$, we have that $M, \nu, t \models \varphi$.⁷

In the following, we present a few examples of well-formed sentences that can be written in ISL. Note, however, that only one of them is generally valid (i.e. true in all models), while the others can be considered true in more specific scenarios where the geometry of objects and possible movements are further restricted in the description of the semantics. Alternatively, ISL theories can be used to prescribe admissible spatiotemporal models.

- $FrontOf(a, b) \wedge \mathbf{F}\neg FrontOf(a, b) \longrightarrow \mathbf{F}(a \rightsquigarrow b \vee a \leftarrow b \vee b \rightsquigarrow a \vee b \leftarrow a)$. ‘If a is in front of b , but ceases to be so in the future, then sometime in the future, either a or b must move with respect to the other object’s original position;’
- $Above(a, b) \wedge \mathbf{G}a \circ b \longrightarrow \mathbf{G}Above(a, b)$. ‘If a is above b and never moves relative to b , it will be always above b .’ This sentence is not valid: consider e.g. that a circles around b with constant distance. However, it holds if for example a and b always stay on the same line (that is, their relative movement is 1D only);
- $DC(a, b) \wedge \mathbf{G}a \leftarrow b \longrightarrow \mathbf{G}DC(a, b)$. ‘If a is disconnected from b , and always moves away from it, it will always stay disconnected from b .’ It can be seen that this formula is, in fact, a validity.

4 Three Types of Image Schema Combinations

Formalising image schemas using ISL makes it possible to represent the individual image schemas. Additionally, by taking their spatial, and temporal, primitives (such as PATH, OBJECT, OUTSIDE and INSIDE [49]) into account, similar image schemas can be grouped together into ‘families’ represented as graphs of theories with increasing complexity [33]. The latter provides a means to investigate the merged combinations of image schemas by looking at the intersection of two different image schema families (i.e. ‘Going IN’ would lie at the intersection of SOURCE_PATH_GOAL and CONTAINMENT). The collection of formalised image schemas and their spatial components can be seen as a repository of cognitively-based *ontology design patterns* [16] that can be used when building conceptualisations of

⁷ Note that, in order to keep the semantics simpler and within first-order logic, only quantification over rigid objects are allowed.

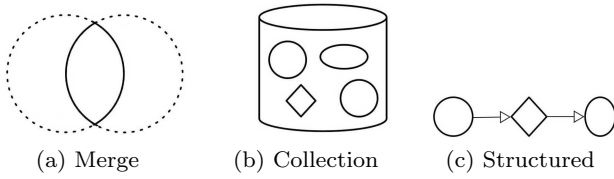


Fig. 1: Three different ways image schemas can be combined with each other.

concepts and events. In the next section, we illustrate this phenomenon by generating image schema profiles for Egg Cracking.

We argue that image schema combinations come in (at least) three fundamentally different flavours. The basic intuition behind these combination approaches is illustrated in Fig. 1. To briefly summarise the three approaches, assume a ‘small’ finite set of atomic image schemas \mathfrak{A} is given, namely those that are cognitively learned first and cannot be further decomposed.

Firstly, the **merge** operation takes a number of those image schemas and merges them (non commutatively) into newly created primitive concepts. These primitives are not yet logically analysed, but carry strong cognitive semantics. This process can be iterated to create ever more complex primitives, as happens in the cognitive development of children. We provide examples for this procedure below. Therefore, the merge operation multiplies the set of available image schema primitives.

Secondly, the **collection** operation technically corresponds to the formation of an unsorted multiset of atomic and merged image schemas used to describe scenes or objects in a complex scenario, again discussed further below.

Thirdly, **structured** covers the case where, on the one hand, merged image schemas receive a formal semantics, and on the other hand, the temporal interaction that is absent in the ‘collection’ scenario is formally made explicit using temporal logic.

4.1 Merges: Atomic Combinations turn into Complex Image Schemas

Image schemas can be both static and dynamic, meaning that it is possible to add a temporal dimension to many static image schemas; consider for instance the difference between Contained_Inside and Going_IN. However, image schemas are spatiotemporal and it is possible to add or remove *spatial* primitives as well. Building from the hierarchy from [49], where spatial primitives are separated from image schemas and image

schemas separated from conceptual integrations,⁸ Hedblom et al. [33] present the idea that image schemas can be formally organised into families of logical theories, structured hierarchically reflecting increasing complexity by the addition (or removal) of conceptual primitives. This paves the way to address complex image schemas that involve spatial (and temporal) primitives originating from different image schema families. When image schemas are sorted into such graphs, there are intersections where different schema families overlap. For instance, even though Going_IN is often conceptualised as an atomic image schema in its own right, it is arguably better analysed as a SPG that results in an instance of CONTAINMENT (see [31] for a deeper analysis). This in fact gives a good example for the non-commutative nature of the ‘merge’ operation, that we here denote by \mathbb{M} . Given the primitives $s, c \in \mathfrak{A}$ (for SPG and containment), we obtain the merges $s \mathbb{M} c$ and $c \mathbb{M} s$ creating two new primitives that take the sum of the arguments of the component image schemas, but where the first corresponds to Going_IN and the latter to Going_OUT.

Likewise, the more advanced image schema REVOLVING_MOVEMENT is part of the SPG family, yet it can be argued that it inherits the revolving pattern from the image schema CYCLE and the spatial proportions of CENTER_PERIPHERY.

This line of combining image schemas to build new ones can be interpreted as a particular instance of the theory of conceptual blending, introduced in [17]. The theory proposes that all novel ideas are a result of *blending* already existing information by re-combining the given information selectively (see [15] for a formal computational treatment, and [11, 12] for general overviews). Given that blending is a fundamental principle of generation, one of the most basic forms of combining image schemas is, therefore, to selectively blend properties of different image schemas into new ones. For instance, the established image schema LINKED_PATH can be reconstructed as a combination of properties from both SPG and LINK. This merge can be used in the real world in relation to concrete concepts such as trucks with trailers, or in more abstract scenarios such as *marriage* which often is conceptualised as two people walking together through life [47].

We present **merge** here as first combination technique because it operates initially on primitive, and not further de-composable image schemas (and which are typically acquired first also in development), such as containment. It then creates via successive blending the general pool of (complex) image schemas that can

⁸ Conceptual integrations are described as image schemas with the addition of a non-spatial element such as force.

be further used in **collection**, discussed next, and in **structured**.

4.2 Collections: Classic Image Schema Profiles

The second form of image schematic combination, here called **collection**, is where image schemas co-exist to describe a concept, distinct from their own properties. For instance, the concept *transportation* actualises the image schemas SPG and SUPPORT (or CONTAINMENT) [39], but the image schemas themselves are not merged, they are simply grouped together to capture the conceptualisation of the concept; that is, they each provide relevant properties for the overall schema. Experiments have been performed to demonstrate this phenomenon of using image schemas to describe the essence of objects, for instance, [24] and Chapter 7 in [30]. In [55], these profiles are specifically described to be without any particular structure or order. Instead, they are thought to correspond to the gathered experience a person has with a particular concept. For instance, when presented with a familiar scenario, e.g., *going to the supermarket* or *borrowing a book at the library*, we have a mental generalisation based on all previous (explicit and implicit) experiences with that particular scenario and have a mental space for that concept that we use to verbalise our thoughts when conversing and interacting with other people. In the more generic, often-experienced situations, human conceptualisations can be argued to be greatly overlapping across people. For instance, despite strong cultural differences, it is likely that all humans share the same, or essentially indistinguishable, conceptualisation of the concepts of *being hungry* and *going to sleep* as they are fundamentally embodied in their nature. For events such as *going to war* or *preparing Turducken*⁹ which many of us never experience first hand, our conceptualisations are based on the accounts of others. This is one of the strengths of the human mind. Namely, that a person who never cooked *Turducken* can still create an image schema profile to capture the process of preparing the dish. One such conceptualisation could consist of: going IN—as the chicken goes into the duck, and the duck goes into the turkey; CONTAINMENT—as the animals remain inside ‘each other;’ ITERATION—as this process is repeated three times; and SCALE—as the chicken, the duck, and the turkey are treated in their respective sizes. Naturally, an expert chef frequently preparing the dish might understand that there is more at work. This form of combining image schemas behaves like **collec-**

tions as they are without any internal structure and temporal or hierarchical order.

4.3 Structured: Sequential Image Schema Combinations

A metaphorical example for a sequential combination is the idiom *to hit a wall*. In many contexts, this does not mean to physically crash into a wall but instead implies some form of mental or physical breakdown, often preceded by long-term stress or exhausting efforts. The idiom captures the image schema of BLOCKAGE. It is clear that BLOCKAGE is not an atomic image schema but rather a **sequential** combination of several ones (see [6, 32] for in-depth analyses). It would not be inaccurate to describe BLOCKAGE as a merge of other image schemas, as it is built on primitives from several image schema families (among other SPG and CONTACT) but it is more useful to acknowledge the sequential dimension of the image schema; basically, the presence of a cause-and-effect relationship. Breaking BLOCKAGE down, there are at least two OBJECTS, a SPG, and at least one time-point when the two objects are in CONTACT, which results in the hindered movement of the object in motion.

These structured sequences are one way in which the conceptualisation of particular *scenes* and *events* can be formally described. Ontologically speaking, events are manifestations of certain dispositions (capabilities, capacities, affordances, and forces) that map the world from situation to situation [26]. A situation, in turn, is a part of reality that can be understood as a whole (e.g., being married to Mary, sitting on a bench, being inside a duck that is itself inside a turkey). According to [1], a scene involves a (temporal) succession of situations and events involving the objects in the scene. In other words, a scene can be seen as a container for situations. The boundaries of these *containers* are typically defined by a spatiotemporal region, i.e., a scene happens in a continuous interval of time and in a convex region of space [25]. Moreover, they are then objects of a *unitary perception act*. In other words, the main characteristic of a scene is that “it is a whole, from a perceptual point of view” [25], without committing to “specific unity conditions for specifying these wholes.” Finally, as discussed in [1], complex events can be seen as decomposed in a number of more elementary scenes, each of which can be understood as a whole.¹⁰

⁹ A dish prepared through the iterative stuffing of a chicken into a duck, and the duck into a turkey.

¹⁰ In the theory proposed in [25], events can also be ‘carved out’ of complex scenes by selecting a particular focus (e.g., a particular disposition). We, however, leave this aspect out of the present discussion.

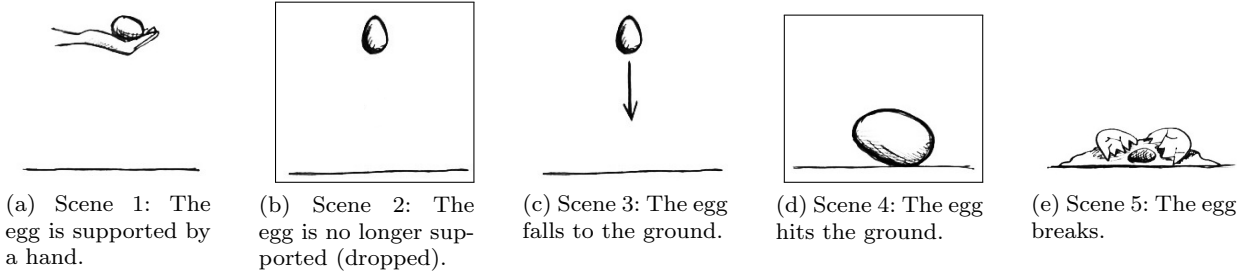


Fig. 2: Event Segmentation of Dropping an Egg. Boxes around scenes denote non-temporally extended scenes which mark essential transitions in image-schematic structure.

$$\begin{aligned}
 \forall E:Obj, H, G:Rgn. & \text{SUPPORT}(H, E) \text{ U } (\neg \text{CONTACT}(H, E) \wedge \text{On_PATH_To}(E, G) \text{ U } \text{BLOCKED}(E, G)) \wedge \\
 & \mathbf{G}(\text{BLOCKED}(E, G) \wedge \mathbf{H}\neg \text{BLOCKED}(E, G) \wedge \mathbf{P}\text{On_PATH_To}(E, G) \rightarrow \text{SPLITTING}(E) \wedge \mathbf{G}\text{SUPPORT}(G, E)) \wedge \\
 & \mathbf{G}(\text{SUPPORT}(H, E) \rightarrow \text{CONTACT}(H, E))
 \end{aligned}$$

Fig. 3: Formalisation of dropping an egg.

The structured sequences of image schemas that we propose here to model events, in a sense, resemble Schankian scripts [60], but with the crucial difference that each scene in the sequence is defined by a potentially different image-schematic structure. This is an important distinction as the image schemas are inherently meaningful and would as such be the core meaning of a particular present situation. Therefore, one could assume that a particular event segment (i.e., a scene) remains the same as long as there is no alteration in the image-schematic structure. In other words, we propose here that image-schematic structures give rise to ‘specific unity conditions’ for individuating scenes. This is properly demonstrated in the egg cracking events presented in Section 5.

For the remainder of this article, we concentrate on formalising this particular mode of image-schema combinations (structured sequential combinations). An important aspect to note here is that, whilst structured image schema profiles may have a clearly determined outcome, in many natural scenarios the outcomes of ongoing and future events are uncertain. This means that also the conceptualisation needs to represent the different possible outcomes of such uncertainty. In the scenario of *BLOCKAGE*, for instance, in which one object moves to collide with a second object, there are several different outcomes (e.g. *CAUSED_MOVEMENT* or *BOUNCES*). This means that structured image schema combinations may also be branching over points of uncertainty.

5 Studies in Egg Cracking with Image Schemas

One of the prototypical knowledge representation problems, ‘cracking an egg,’ is—as an event—rather simple to conceptualise yet very complex to formalise. Previous formalisations of the problem [45, 51] result in lengthy descriptions where individual axioms aim to capture all the necessary requirements for the scenario, with a particular difficulty in formally separating high-level schematic conceptualisation from the formalisation of low-level, physics-based information related to affordances. When taking the embodied point of view which motivates our modelling based on image schemas, such low-level modelling is largely abstracted away. Instead, e.g. the verification of the affordance of an object to contain a liquid is taken care of by embodied interaction in the case of humans, and by experiment in physics simulations in the case of AI (see below for an outlook to future work in this regard). Following the reasoning in this paper, it is possible to use image schema profiles, or more structured image schema combinations, as a way to represent conceptual information. We look at two different scenarios.

5.1 Dropping an Egg

Infants do not have enough experience with the object ‘egg’ to immediately understand that when dropped, eggs fall and as they hit the ground they (usually) break. This knowledge is learned through repeated experience. While temporally dependent scenarios happen in more or less a sequence without defined borders, the event can be divided into conceptually distinct steps

based on changes in the image-schematic structure, as depicted in Fig. 2.

One important hypothesis is that, for each step, a conceptually different scene of undefined temporal length takes place. This translates into there being a change in the image-schematic state. The scenario can be described with a *sequential* image schema combination based on the following scenes.

1. The egg is SUPPORTED by a hand.¹¹
2. The egg is no longer SUPPORTED. In most natural cases there is still CONTACT between the hand and the egg at this stage. In a human conceptualisation, this event takes place more or less simultaneously as the consecutive scene in which ...
3. ...the egg falls from the SOURCE (hand), to the GOAL, where *falling* is a **merge** between SPG and VERTICALITY as the gestalt properties of each image schema rely on one another.
4. The egg is BLOCKED by the ground, stopping its SOURCE_PATH_GOAL.
5. This final scene produces an image-schematic transformation of a SPLITTING in which we observe that $WHOLE(egg) \rightarrow PARTS(egg)$,¹² and the egg remains SUPPORTED by the ground.¹³

As defended in [68], ontology modelling patterns should be construed as generic modelling structures that reflect ontological micro-theories. As such, they constitute a mechanism for *theory inclusion* such that there is a set of generic axioms associated with the pattern structure. Whenever the pattern is reused, so are the corresponding axiomatisations. Primitive patterns can be combined to form larger patterns that consistently preserve this mechanism [16, 68]. This general idea of ontology pattern also underlies the axiomatisation of image schemas in [33] and the use of the Distributed Ontology Language (DOL) that supports exactly this kind of theory inclusion, amongst many other structuring features [52].

The idea of modular design pattern is also reflected in the construction of ISL, where each image schema can be formalised as a modelling pattern, a micro-theory, which can be referenced and reused in different situations and contexts for entirely different kinds of objects

¹¹ It is possible to substitute SUPPORT for CONTAINMENT if the egg is ‘grabbed.’ This would alter the properties of the agent’s involvement in the ‘drop.’

¹² In ISL DC means DisConnected (based on RCC8) and \mathbf{U} , is taken from RTL and denotes ‘Until.’ Thus, the image schema SPLITTING can be formalised as:

$\forall X, x_1, x_2 : \text{Object}(\text{SPLITTING}(x) \rightarrow \text{WHOLE}(X) \wedge \text{Part}(x_1, X) \wedge \text{Part}(x_2, X) \mathbf{U} \neg \text{WHOLE}(X) \wedge \text{DC}(x_1, x_2))$.

¹³ This is due to the BLOCKAGE relation from the previous scene, only now, the force is removed resulting in the SUPPORT (see [32] for details on their respective formalisations).

via a generic import interface (as the case of SPLITTING before). A large selection of these image schema patterns appears in [30]. We limit our formalisation to capture the patterns of the top-level of the event structure, and only report on a few specific image schema micro-theories.

Even when using specifically RCC8 to model image schemas like CONTACT and SUPPORT (and in fact other spatial frameworks could be substituted instead), the most appropriate definition depends on the kinds of objects we consider, and the chosen granularity of observation, amongst other factors. For instance, we may identify contact with external connection when an idealised geometric representation of objects can be assumed.

$$\forall O_1, O_2: \text{Object} (\text{CONTACT}(O_1, O_2) \leftrightarrow \text{EC}(O_1, O_2))$$

However, a much more liberal interpretation of contact is obtained when we identify contact with the absence of disconnectedness, as in:

$$\forall O_1, O_2: \text{Object} (\text{CONTACT}(O_1, O_2) \leftrightarrow \neg \text{DC}(O_1, O_2))$$

Several intermediate options are obviously available as well.¹⁴

In ISL the entire event of *dropping an egg* could be formalised as in Fig. 3, where E , H , and G stand for *Egg*, *Hand*, and *Ground*, respectively. Note that the figure has two ontologically quite distinct kinds of scenes. Namely, whilst (a), (c), and (e) describe temporally extended scenes, (b) and (d) describe idealised moments that mark the transition between the respective frames. Importantly, the image schema profiles of all scenes are distinct (in particular (d) has different image schemas related to force compared to (a) as it follows vertical movement).

The axiom given specifies the following: the first line encodes the global event structure, namely that the hand provides support to the egg, until it is no longer in contact with it, at which point it will be on its way towards the ground. This scene will last until the egg will be blocked by the ground, at an unknown point in the future. The remaining axiomatisation encodes some of the essential properties that need to hold for this particular outcome, and that are part of the commonsensical understanding of ‘dropping.’ Namely, the second line says that if at any point in the future the egg is blocked by the ground, but it was never blocked before but instead was at some point in the past moving towards the ground, then it will now break, and it

¹⁴ Similarly, we may choose to define SUPPORT as $\forall O_1, O_2: \text{Object} (\text{SUPPORT}(O_1, O_2) \leftrightarrow \text{EC}(O_1, O_2) \wedge \text{Above}(O_1, O_2) \wedge \text{Forces}(O_1, O_2))$, etc.

will be then supported forever by the ground. Finally, the last line encodes that the hand gives support to the egg only if it is in contact with the egg.

5.2 Cracking an Egg into a Bowl

In most scenarios where there is an intention to crack the egg, this is done by gathering the contents in a bowl. Such an event can be divided into ten conceptually distinct spatiotemporal scenes, as depicted in Fig. 4. Note that, as above, the schematicity of the description implies that the more detailed axiomatisations (or indeed other ways of grounding the truth of those predicates) is left to the refinement of the schema. For example, when saying that an egg can be seen as a whole with ‘parts,’ ‘inside,’ and ‘content,’ it is at this level of description left open what the exact definition of whole as mereological sum of its parts is.

1. Scene one presupposes two OBJECTS: an egg and a bowl. The bowl is a CONTAINER and represents the egg’s GOAL location. Additionally, the egg needs to be described as a WHOLE with two PARTS: the shell (CONTAINER) and an egg’¹⁵ (CONTAINED). This is a conceptual **merge** between CONTAINMENT and PART_WHOLE.¹⁶
2. Scene two extends scene one with a SPG as the egg is moving from its original position towards the edge of the bowl.
3. As the egg hits the border of the bowl, the movement is BLOCKED. This means that instead of the previous SPG image schema, the image-schematic relationship is that of BLOCKAGE. As the egg hits the edge of the bowl, it is intended to crack. However, conceptually this is a different event component that may or may not take place, depending on the characteristics of the impact between the bowl and the egg. Then ...
4. ...the egg cracks: breaking from a WHOLE into its PARTS: the shell and the egg’. This is an image-schematic transformation of PART_WHOLE. While this event may be perceived to happen simultaneously as the third scene, it is conceptually different because the properties of the egg suddenly are altered. Likewise, if insufficient force is applied there

¹⁵ While in natural language both the whole egg and its content is referred to as an egg, we need to formally distinguish them. Thus, we refer to the whole egg as *egg* and the content as *egg’*.

¹⁶ For eggs, it is rather straightforward that the part that we use is on the inside of the shell. However, consider an apple or other objects in which the ‘border’ is (most often) used as well. In these cases it is not appropriate to speak of a merge between CONTAINMENT and PART_WHOLE in the same sense.

is no guarantee that the egg cracks or if excessive force is applied the egg’ pours out all over the bowl’s edge (considerations on *force* are addressed in Section 5.3).

5. Still CONTAINED in the cracked shell, the egg’ moves towards the bowl’s opening. This scene functions as a **collection** (neither is dependent on the other) and captures both CONTAINMENT and SPG.
6. Removing the CONTAINMENT schema of the egg, by SPLITTING the shell from the egg’ through the existence of their PART_WHOLE relationship.
7. As a **merge**, the egg’ goes OUT from the shell and begins to fall towards the bowl’s INSIDE.
8. The egg’ continues to fall towards the bowl’s inside.
9. Still moving, the egg’ falls into the bowl: the **merge** between Going-IN and the pre-existing **merge** of falling based on SPG and VERTICALITY.
10. Finally, the scenario ends with static CONTAINMENT in which the egg’ rests inside the bowl.

A formalisation appears in Fig. 5, where *E, E’, B, H, S* stand for *Egg, Egg’, Bowl, Hand, and Shell*, respectively. The detailed semantics of this can be recovered as in the previous example.

5.3 The Problem of Force in Egg Cracking

One of the limitations of the egg cracking scenarios presented is that they both represent the ideal ‘successful’ scenario. For an egg falling to the ground, the most natural outcome is that it will hit the ground and break. In an unsuccessful scenario, the egg might not actually break. This could be the result of an unusually hard shell, a ‘soft landing’ on a carpet or that it has been dropped from a low height. All of this comes down to one physical component, that of *force*.

Image schemas have several force relations built into them. For instance, SUPPORT relies on the notion that enough force keeps the object in place, and BLOCKAGE captures the counterforce equivalent (or stronger) present in the movement. In [49], the authors describe the concept of force as an embodied, conceptual add-on to image schemas. When modelling a given scenario, propositional add-ons such as the hardness of the shell or the ground, the height of the drop or the force by which the egg hits the bowl can be attached to the image-schematic skeleton of the individual scenario to provide a more detailed description. A cognitively inspired approach to detect whether ‘enough’ aspects describing a certain scenario or concept are accumulated in a concrete modelling was introduced in [57].

However, as already hinted at in the introduction, image-schema-level formalisations are not intended to

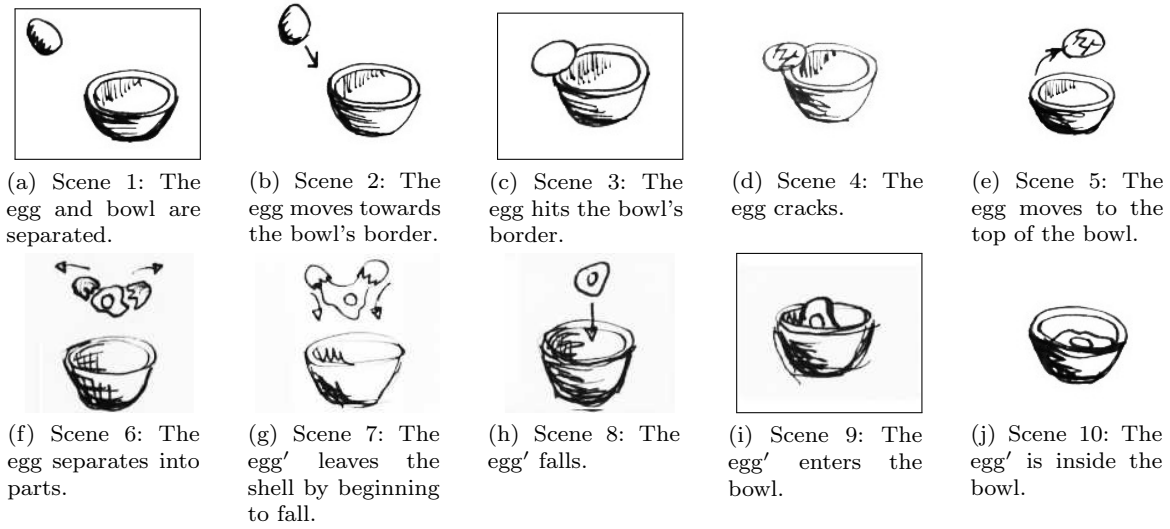


Fig. 4: Event Segmentation of Cracking an Egg into a Bowl. Boxes denote the same distinction as previously.

$$\begin{aligned}
 &\forall E, E', B: \text{Object}, H, S: \text{Region}. \text{Contained_Inside}(H, E) \wedge \text{WHOLE}(E) \wedge \text{PART}(S, E) \wedge \text{PART}(E', E) \wedge \text{Contained_Inside}(E', S) \wedge \\
 &\quad \mathbf{F}(\text{On_PATH_To}(E, B) \mathbf{U} (\text{BLOCKAGE}(\text{Egg}, \text{Bowl}) \wedge \mathbf{FSPLITTING}(\text{Egg}))) \wedge \\
 &\quad \mathbf{G}(\text{SPLITTING}(E) \rightarrow \\
 &\quad \quad (\mathbf{G} \neg \text{WHOLE}(E) \wedge \text{going_OUT}(S, E') \wedge \text{On_PATH_To}(E', B) \mathbf{U} \text{going_IN}(E', B))) \wedge \\
 &\quad \mathbf{G}(\text{going_IN}(E', B) \rightarrow \mathbf{F}\text{Contained_Inside}(E', B))
 \end{aligned}$$

Fig. 5: Formalisation of cracking an egg in a bowl.

cover the low-level physics of a scenario. Rather, the force dynamic events that can be detected in, e.g., the physics simulations of robotics environments can trigger image-schematic primitives without a logical analysis of causation and force [27]. Therefore, the *actual* outcome of an open-ended formalisation of an everyday scenario such as ‘cracking an egg’ can only be determined if the precise force acting on the egg is known, and this can be read off the virtual enactment of the egg hitting the bowl in a simulation with precise physics.

6 Discussion and Conclusions

This paper studies how image schema combinations can be structured and formally approached to model the conceptualisation of dynamic concepts and events. In particular, event segmentation into ontologically and cognitively meaningful scenes can be based on changes in image-schematic state and modelled as a structured combination of component scenes. To this end, we introduce three different categories for the combination of image schemas: merge, collection and structured. The first captures the proliferation of image-schematic primitives, the second the collection of those primitives into

new wholes, and the third the temporal arrangement of collections. While these forms of combinations capture some of the most apparent combinations of image schemas, they are by no means intended to be exhaustive. Other combinations, or even combinations of these combinations, which were not considered in this paper, may be worthwhile to study in future work. The image schemas within these profiles were then formalised using ISL, a logical language especially developed to deal with the spatiotemporal dimensions of image schemas.

Arguably, looking at commonsense reasoning problems such as egg cracking may look a bit isolated in terms of their potential impact on artificial intelligence. However, the idea of using cognitively-inspired building blocks that can together represent and model increasingly large-scale situations and problems is of wide relevance. As the notion of image schemas stems from the sensorimotor processes and is closely connected to cognitive linguistics, their formal integration into robotics systems and natural language processing systems provides clear directions for future work. Indeed, the next step on this research agenda is to connect our approach to cognitive robotics environments as for instance described in [64]. Here, symbols may be grounded in ac-

tual environments, and symbolic twin-worlds and knowledge bases, together with physics simulations, can provide precise tests for preconditions of actions and events whose detail, for instance, in the level of force present, escapes the image-schematic modelling level.

References

1. J. P. A. Almeida, P. D. Costa, and G. Guizzardi. Towards an ontology of scenes and situations. In *2018 IEEE Conf. on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, pages 29–35. IEEE, 2018.
2. D. A. Baldwin, J. A. Baird, M. M. Saylor, and M. A. Clark. Infants parse dynamic action. *Child development*, 72(3):708–717, 2001.
3. J. A. Bateman, J. Hois, R. Ross, and T. Tenbrink. A Linguistic Ontology of Space for Natural Language Processing. *Artificial Intelligence*, 174(14):1027–1071, 2010.
4. N. Belnap, M. Perloff, and M. Xu. *Facing the Future (Agents and Choices in Our Indeterminist World)*. Oxford University Press, 2001.
5. B. Bennett and C. Cialone. Corpus Guided Sense Cluster Analysis: a methodology for ontology development (with examples from the spatial domain). In P. Garbacz and O. Kutz, editors, *8th Int. Conf. on Formal Ontology in Inform. Systems (FOIS)*, volume 267, pages 213–226. IOS Press, 2014.
6. T. R. Besold, M. M. Hedblom, and O. Kutz. A narrative in three acts: Using combinations of image schemas to model events. *Biologically Inspired Cognitive Architectures*, 19:10–20, 2017.
7. I. Biederman. Recognition by components: A theory of human image understanding. *Psychological Review*, 94(2):115–117, 1987.
8. L. Boroditsky. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75(1):1–28, 2000.
9. A. Cienki. Some properties and groupings of image schemas. In M. Verspoor, K. D. Lee, and E. Sweetser, editors, *Lexical and Syntactical Constructions and the Construction of Meaning*, pages 3–15. John Benjamins Publishing Company, Philadelphia, 1997.
10. A. G. Cohn, B. Bennett, J. Gooday, and N. Gotts. RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, 1:275–316, 1997.
11. R. Confalonieri and O. Kutz. Blending under Deconstruction: The roles of logic, ontology, and cognition in computational concept invention. *Annals of Mathematics and Artificial Intelligence*, 2019. To appear.
12. R. Confalonieri, A. Pease, M. Schorlemmer, T. Besold, O. Kutz, E. Maclean, and M. Kaliakatsos-Papakostas, editors. *Concept Invention: Foundations, Implementation, Social Aspects and Applications*. Computational Synthesis and Creative Systems. Springer, 2018.
13. S. Coulson and C. P. Cánovas. Understanding Timelines: Conceptual Metaphor and Conceptual Integration. *Cognitive Semiotics*, 5(1-2):198–219, 2014.
14. E. Dodge and G. Lakoff. Image schemas: From linguistic analysis to neural grounding. In B. Hampe and J. E. Grady, editors, *From perception to meaning: Image schemas in cognitive linguistics*, pages 57–91. Mouton de Gruyter, Berlin, New York, 2005.
15. M. Eppe, E. Maclean, R. Confalonieri, O. Kutz, M. Schorlemmer, E. Plaza, and K.-U. Kühnberger. A computational framework for conceptual blending. *Artificial Intelligence*, 256:105–129, 2018.
16. R. Falbo, G. Guizzardi, A. Gangemi, and V. Presutti. Ontology patterns: Clarifying concepts and terminology. In *Proc. 4th Workshop on Ontology and Semantic Web Patterns*, 2013.
17. G. Fauconnier and M. Turner. Conceptual integration networks. *Cognitive Science*, 22(2):133–187, 1998.
18. M. Finger and D. M. Gabbay. Adding a Temporal Dimension to a Logic System. *Journal of Logic, Language and Information*, 1:203–233, 1993.
19. A. U. Frank and M. Raubal. Formal specification of image schemata – a step towards interoperability in geographic information systems. *Spatial Cognition and Computation*, 1(1):67–101, 1999.
20. A. Galton. Some problems in temporal knowledge representation. In *Logic and Change, GWAI-92*, Bonn, Germany, September 1992.
21. A. Galton. The Formalities of Affordance. In M. Bhatt, H. Guesgen, and S. Hazarika, editors, *Proc. of the workshop Spatio-Temporal Dynamics*, pages 1–6, 2010.
22. R. W. Gibbs, D. A. Beitel, M. Harrington, and P. E. Sanders. Taking a stand on the meanings of stand: Bodily experience as motivation for polysemy. *Journal of Semantics*, 11(4):231–251, 1994.
23. J. A. Goguen and R. M. Burstall. Institutions: Abstract Model Theory for Specification and Programming. *Journal of the Association for Computing Machinery*, 39(1):95–146, 1992. Predecessor in: LNCS 164, 221–256, 1984.
24. D. Gromann and J. C. Macbeth. Crowdsourcing image schemas. In *Proc. of TriCoLore*, volume 2347 of *CEUR-WS*, Bolzano, Italy, 2019.

25. N. Guarino and G. Guizzardi. Relationships and events: Towards a general theory of reification and truthmaking. In *Proc. of XVth Intern. Conf. of the Italian Association for Artif. Intel. (AI*IA 2016)*, pages 237–249, 2016.
26. G. Guizzardi, G. Wagner, R. de Almeida Falbo, R. S. S. Guizzardi, and J. P. A. Almeida. Towards ontological foundations for the conceptual modeling of events. In *Proc. of 32nd Intern. Conf. Conceptual Modeling (ER 2013)*, pages 327–341, 2013.
27. A. Haidu, D. Beßler, A. K. Bozcuoglu, and M. Beetz. KNOWROB-SIM: Game Engine-enabled Knowledge Processing for Cognition-enabled Robot Control. In *International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, 2018. IEEE.
28. B. Hampe and J. E. Grady. *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*. Walter de Gruyter, Berlin, 2005.
29. P. J. Hayes. The naive physics manifesto. In *Band 34 von Fondazione Dalle Molle per gli Studi Linguistici e di Comunicazione Internazionale*. University of Essex, 1978.
30. M. M. Hedblom. *Image Schemas and Concept Invention: Cognitive, Logical, and Linguistic Investigations*. PhD thesis, Otto-von-Guericke University of Magdeburg, 2018.
31. M. M. Hedblom, D. Gromann, and O. Kutz. In, Out and Through: Formalising some dynamic aspects of the image schema Containment. In *Proc. 33rd Annual ACM Symposium on Applied Computing*, pages 918–925, Pau, France, 2018.
32. M. M. Hedblom, O. Kutz, T. Mossakowski, and F. Neuhaus. Between contact and support: Introducing a logic for image schemas and directed movement. In F. Esposito, R. Basili, S. Ferilli, and F. A. Lisi, editors, *AI*IA 2017: Advances in Artificial Intelligence*, pages 256–268, 2017.
33. M. M. Hedblom, O. Kutz, and F. Neuhaus. Choosing the Right Path: Image Schema Theory as a Foundation for Concept Invention. *Journal of Artificial General Intelligence*, 6(1):22–54, 2015.
34. M. M. Hedblom, O. Kutz, R. Peñaloza, and G. Guizzardi. What’s Cracking: How image schema combinations can model conceptualisations of events. In *Proc. of TriCoLore*, volume 2347 of *CEUR-WS*, Bolzano, Italy, 2019.
35. M. Johnson. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press, 1987.
36. M. Knauff, R. Rauh, and J. Renz. A cognitive assessment of topological spatial relations: Results from an empirical investigation. In S. C. Hirtle and A. U. Frank, editors, *Spatial Information Theory: A Theoretical Basis for GIS*, volume 1329 of *Lecture Notes in Computer Science*, pages 193–206. Springer, 1997.
37. M. Kracht and O. Kutz. Logically Possible Worlds and Counterpart Semantics for Modal Logic. In D. Jacquette, editor, *Philosophy of Logic, Handbook of the Philosophy of Science*, volume 5, pages 943–996. Elsevier, Amsterdam, 2007.
38. F. Kröger and S. Merz. *Temporal Logic and State Systems (Texts in Theoretical Computer Science. An EATCS Series)*. Springer, first edition, 2008.
39. W. Kuhn. An Image-Schematic Account of Spatial Categories. In S. Winter, M. Duckham, L. Kulik, and B. Kuipers, editors, *Spatial Information Theory*, volume 4736 of *LNCS*, pages 152–168. Springer, 2007.
40. C. A. Kurby and J. M. Zacks. Segmentation in the perception and memory of events. *Trends in cognitive sciences*, 12(2):72–79, 2008.
41. O. Kutz, C. Lutz, F. Wolter, and M. Zakharyashev. \mathcal{E} -Connections of Abstract Description Systems. *Artificial Intelligence*, 156(1):1–73, 2004.
42. O. Kutz, T. Mossakowski, and D. Lücke. Carnap, Goguen, and the Hyperontologies: Logical Pluralism and Heterogeneous Structuring in Ontology Design. *Logica Universalis*, 4(2):255–333, 2010. Special Issue on ‘Is Logic Universal?’.
43. O. Kutz, N. Troquard, M. M. Hedblom, and D. Porello. The Mouse and the Ball: Towards a Cognitively-Based and Ontologically-Grounded Logic of Agency. In *Proceedings of the 10th International Conference on Formal Ontology in Information Systems (FOIS 2018)*, 2018.
44. G. Lakoff. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. University of Chicago Press, 1987.
45. V. Lifschitz. Cracking an egg: An exercise in commonsense reasoning. In *Proc. of the Symp. of Logical Formalizations of Commonsense Reasoning*, 1998.
46. G. Ligozat. Reasoning about cardinal directions. *J. Vis. Lang. Comput.*, 9(1):23–44, 1998.
47. J. M. Mandler. *The Foundations of Mind: Origins of Conceptual Thought: Origins of Conceptual Thought*. Oxford University Press, New York, 2004.
48. J. M. Mandler. Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development*, 1(1):3–36, 2009.
49. J. M. Mandler and C. Pagán Cánovas. On defining image schemas. *Language and Cognition*, 6(4):510–532, may 2014.

50. N. Markey. Past is for free: On the complexity of verifying linear temporal properties with past. *Acta Inf.*, 40(6-7):431–458, May 2004.
51. L. Morgenstern. Mid-Sized Axiomatizations of Commonsense Problems: A Case Study in Egg Cracking. *Studia Logica*, 67:333–384, 2001.
52. T. Mossakowski, M. Codescu, F. Neuhaus, and O. Kutz. *The Road to Universal Logic—Festschrift for 50th birthday of Jean-Yves Beziau, Volume II*, chapter The distributed ontology, modelling and specification language - DOL. Studies in Universal Logic. Birkhäuser, 2015.
53. T. Mossakowski, R. Diaconescu, and A. Tarlecki. What is a logic translation? *Logica Universalis*, 3(1):95–124, 2009.
54. T. Mossakowski, J. Goguen, R. Diaconescu, and A. Tarlecki. What is a logic? In J.-Y. Beziau, editor, *Logica Universalis*, pages 113–133. Birkhäuser, 2005.
55. T. Oakley. Image schema. In D. Geeraerts and H. Cuyckens, editors, *The Oxford Handbook of Cognitive Linguistics*, pages 214–235. Oxford University Press, 2010.
56. C. Pagán Cánovas. Erotic Emissions in Greek Poetry: A Generic Integration Network. *Cognitive Semiotics*, 6:7–32, 2010.
57. D. Porello, O. Kutz, G. Righetti, N. Troquard, P. Galliani, and C. Masolo. A Toothful of Concepts: Towards a theory of weighted concept combination. In M. Šimkus and G. Weddell, editors, *Proceedings of the 32th International Workshop on Description Logics (DL 2019)*, volume 2373, Oslo, Norway, June 18–21, 2019. CEUR-WS.
58. D. A. Randell, Z. Cui, and A. G. Cohn. A spatial logic based on regions and connection. In *Proc. 3rd Int. Conf. on Knowledge Rep. and Reas.*, 1992.
59. M. Reynolds. The complexity of temporal logic over the reals. *Annals of Pure and Applied Logic*, 161(8):1063 – 1096, 2010.
60. R. C. Schank and R. P. Abelson. Scripts, plans, and knowledge. In *IJCAI*, pages 151–157, 1975.
61. C. D. Schunn and A. H. Vera. Causality and the categorisation of objects and events. *Thinking & Reasoning*, 1(3):237–284, 1995.
62. A. Scivos and B. Nebel. The Finest of its Class: The Natural, Point-Based Ternary Calculus \mathcal{LR} for Qualitative Spatial Reasoning. In *Spatial Cognition*, pages 283–303, 2004.
63. L. Talmy. The fundamental system of spatial schemas in language. In B. Hampe and J. E. Grady, editors, *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 of *Cognitive Linguistics Research*, pages 199–234. Walter de Gruyter, 2005.
64. M. Tenorth and M. Beetz. Knowrob: A knowledge processing infrastructure for cognition-enabled robots. *The International Journal of Robotics Research*, 32(5):566–590, 2013.
65. N. Van De Weghe, A. G. Cohn, G. De Tré, and P. D. Maeyer. A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and cybernetics*, 35(1):97–119, 2006.
66. A. Wierzbicka. *Semantics: Primes and Universals*. Oxford University Press, UK, 1996.
67. J. M. Zacks, N. K. Speer, K. M. Swallow, T. S. Braver, and J. R. Reynolds. Event perception: a mind-brain perspective. *Psychological bulletin*, 133(2):273, 2007.
68. E. Zambon and G. Guizzardi. Formal definition of a general ontology pattern language using a graph grammar. In *Federated Conference on Computer Science and Information Systems (FedCSIS 2017)*, pages 1–10. IEEE, 2017.