Hold—Drop—Smash
Image Schema Combinations and Complex Events*

Maria M. Hedblom[1], Oliver Kutz[1], Rafael Peñaloza[2], and Giancarlo Guizzardi[1]

[1] Conceptual and Cognitive Modelling Research Group CORE, KRDB Research Centre for Knowledge and Data, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy {mhedblom,okutz,gguizzardi}@unibz.it
[2] Department of Computer Science, Systems and Communications, University of Milano-Bicocca, Italy
rafael.penaloza@unimib.it

Extended Abstract

Motivation

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. Inspired by this, previous representations of the cognitive perception of real world scenes were sometimes based on formal frameworks used in naïve physics [10], such as situation calculus or causal logic. Through these, classic commonsense reasoning problems such as *cracking an egg* [20] were then often described with long and complex axiomatisations that offer little in terms of cognitive adequacy or conceptual clarity.

Formally it is an important distinction that, unlike for objects, there are no 'borders' in the passing of time. One event often floats seamlessly into another without pauses, beginnings or ends. The human mind has an ability to take dynamic perceptions and, based on certain cognitive principles grounded in spatiotemporality, identify when a new *event* takes place [16]. If these principles can be integrated into a formal framework representing events, a more cognitively plausible method for commonsense reasoning over events is possible.

To do this, we look at *image schemas*. They are abstract generalisations of events learned from repeated experiences in the world [18, 15] and are commonly described as capturing sensorimotor patterns of relationships and their transformations. As such, they 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) [5]. 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. [3]). However, in order to represent events and more dynamic concepts, also the temporal and transformational dimension of image schemas require attention. Some work has been done to model the dynamic aspects of image schemas but they are often limited to a

---

* The full paper is published as [14].

particular schema or situation that cannot be easily generalised (e.g. [12, 8]). 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 [21].

We argue that, by using the formal representation of these conceptual primitives in different combinations, it is possible to approach a more cognitively plausible representation of events. Additionally, we argue that change in image-schematic state corresponds to plausible points of division in formal event segmentation. 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 [13]. 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.

Brief Introduction to ISL: The Image Schema Logic

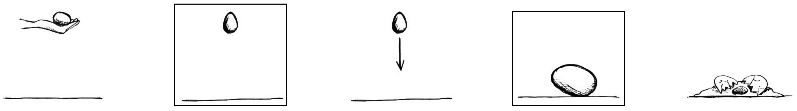
Image schemas are abstract patterns that become detectable only due to their prevalence in natural language and cognition in general. Therefore, much like with all spatiotemporal formalisation problems, it is not trivial to formally represent them in a satisfactory way [9, 1]. One problem for formalising image schemas is that the cognitively-driven investigations of how humans perceive and experience time cannot easily be mapped to existing temporal logic approaches [4]. 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) [6].

While image schemas are often discussed without an immediate formal correspondence, there exists a number of attempts to capture them formally (e.g. [3]). The formal language ISL [13][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 [22], Ligozat's Cardinal Directions (CD) [19], Qualitative Trajectory Calculus (QTC) [24], 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 [12] and expanded to include agency in [17] through the addition of see-to-it-that (STIT) logic [2].

At its core, ISL follows a popular temporalisation strategy (studied in further detail in [7]), 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

---

[3] ISL was further developed under the name $\mathsf{ISL}^{\mathrm{FOL}}$ by the addition of a First-Order concept language in [11].



(a) Scene 1: The egg is supported by a hand.

(b) Scene 2: The egg is no longer supported (dropped).

(c) Scene 3: The egg falls to the ground.

(d) Scene 4: The egg hits the ground.

(e) Scene 5: The egg breaks.

Fig. 1: Event Segmentation of Dropping an Egg. Boxes around scenes denote non-temporally extended scenes which mark transitions in image-schematic structure.

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.

Three Types of Image Schema Combinations

Formalising image schemas using ISL makes it possible to represent the individual image schemas as well as their combinations, a necessary requirement to represent events. We argue that image schema combinations come in (at least) three fundamentally different flavours. 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. 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.

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.

An Egg-Cracking Example

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 (e.g. [20]) result in lengthy descriptions

$$\begin{aligned}\forall E{:}Obj, H, G{:}Rgn.\textsc{Support}(H,E)\ \mathbf{U}\ (\neg\textsc{Contact}(H,E) \wedge \textsc{On\_Path\_To}(E,G) \\ \mathbf{U}\ \textsc{Blocked}(E,G) \wedge \mathbf{G}\,(\textsc{Blocked}(E,G) \wedge \\ \mathbf{H}\neg\textsc{Blocked}(E,G) \wedge \mathbf{P}\textsc{On\_Path\_To}(E,G) \rightarrow \\ \textsc{Splitting}(E) \wedge \mathbf{G}\textsc{Support}(G,E) \wedge \\ \mathbf{G}(\textsc{Support}(H,E) \rightarrow \textsc{Contact}(H,E))\end{aligned}$$

Fig. 2: ISL formalisation of dropping an egg.

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.

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 scenes presented in Figure 1 and the ISL formalisation in Figure 2.

This follows the idea of modular design pattern [25], 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 via a generic import interface. A large selection of these image schema patterns appears in [11]. In ISL, the entire event of *dropping an egg* could be formalised as in Fig. 2, where E, H, and G stand for *Egg*, *Hand*, and *Ground*, respectively. 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).

Although looking at commonsense reasoning problems such as ‘egg cracking’ may look a bit isolated in terms of broader AI research trends, the idea of using cognitively-inspired building blocks that can together represent and model increasingly large-scale situations and problems is in fact 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 [23]. Here, symbols may be grounded in actual 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.

Bibliography

[1] J. A. Bateman. Ontological diversity: the case from space. In A. Galton and R. Mizoguchi, editors, *Formal Ontology in Information Systems - Proceedings of the Sixth International Conference (FOIS 2010)*, volume 209. IOS Press, 2010.

[2] N. Belnap, M. Perloff, and M. Xu. *Facing the Future (Agents and Choices in Our Indeterminist World).* Oxford University Press, 2001.

[3] 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.

[4] L. Boroditsky. Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, 75(1):1–28, 2000.

[5] 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.

[6] A. G. Cohn, B. Bennett, J. Gooday, and N. Gotts. RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, 1:275–316, 1997.

[7] M. Finger and D. M. Gabbay. Adding a Temporal Dimension to a Logic System. *Journal of Logic, Language and Information*, 1:203–233, 1993.

[8] 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.

[9] A. P. Galton. Space, time and movement. In O. Stock, editor, *Spatial and Temporal Reasoning*, chapter 10, pages 321–352. Kluwer, Dordrecht, 1997.

[10] 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.

[11] M. M. Hedblom. *Image Schemas and Concept Invention: Cognitive, Logical, and Linguistic Investigations.* PhD thesis, Otto-von-Guericke University of Madgeburg, 2018.

[12] 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.

[13] 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.

[14] M. M. Hedblom, O. Kutz, R. Peñaloza, and G. Guizzardi. Image schema combinations and complex events. *Knstliche Intelligenz*, Special issue on Cognitive Reasoning, 2019.

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

[16] C. A. Kurby and J. M. Zacks. Segmentation in the perception and memory of events. *Trends in cognitive sciences*, 12(2):72–79, 2008.

[17] 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.

[18] G. Lakoff. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. University of Chicago Press, 1987.

[19] G. Ligozat. Reasoning about cardinal directions. *J. Vis. Lang. Comput.*, 9(1):23–44, 1998.

[20] L. Morgenstern. Mid-Sized Axiomatizations of Commonsense Problems: A Case Study in Egg Cracking. *Studia Logica*, 67:333–384, 2001.

[21] T. Oakley. Image schema. In D. Geeraerts and H. Cuyckens, editors, *The Oxford Handbook of Cognitive Linguistics*, pages 214–235. Oxford University Press, 2010.

[22] 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.

[23] 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.

[24] 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.

[25] 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.