Formal Ontology in Information Systems
B. Brodaric and F. Neuhaus (Eds.)

doi:10.3233/FAIA200668

Embodied Functional Relations: A Formal Account Combining Abstract Logical Theory with Grounding in Simulation

Mihai POMARLAN and John A. BATEMAN
Bremen University, Bremen, Germany

Abstract. Functional relations such as containment or support have proven difficult to formalize. Although previous efforts have attempted this using hybrids of several theories, from mereology to temporal logic, we find that such purely symbolic approaches do not account for the embodied nature of functional relations, i.e. that they are used by embodied agents to describe fragments of a physical world. We propose a formalism that combines descriptions of a high level of abstraction with generative models that can be used to instantiate or recognize arrangements of objects and trajectories conforming to qualitative descriptions. The formalism gives an account of how a qualitative description of a scene or arrangement of objects can be converted into a quantitative description amenable to simulation, and how simulation results can be qualitatively interpreted. We use this to describe functional relations between objects in terms of spatial arrangements, expectations on behavior, and counterfactual expectations for when one of the participants is absent. Our method is able to tackle important questions facing an agent operating in the world, such as what would happen if an arrangement of objects is created and why. This gives the agent a deeper understanding of functional relations, including what role background objects, not explicitly asserted to participate in a functional relation such as containment, play in enabling or hindering the relation from holding.

Keywords. image schemas, embodiment, simulation, linguistic semantics, ontological analysis, formal ontology

1. Introduction: background and motivations

It is a fact universally acknowledged that an agent acting in a world must be in need of understanding how that world works. In humans, such an understanding is grouped under the label of commonsense, which includes aspects of intuitive physics and social behavior, and is acquired through one's own or observed experience, sometimes explicitly taught, but always sinking to an intuitive level that is hard to make explicit again. This presents a challenge for the creation of artificial agents that would be able to operate in the physical world in shared environments with humans. Commonsense turns out to be very difficult to capture formally, and finding the right balance of expressive power versus tractability is an unresolved problem, even despite some recent successes of machine learning in other AI domains. Perennial questions such as the utility of formal ontologies for capturing commonsense everyday knowledge also remain open.

Viewed abstractly, state of the art robot programming consists of following a script which may contain branches for some failure situations. However, problems occur as soon as events outside of the provided script occur. Consider, as an illustrative example, the situation where a robot is tasked with making popcorn. A pot containing oil and some corn kernels lies on a hot stove. Having grabbed a lid, the robot attempts to place it on the pot, but drops it on the way. Unaware of the mistake, the robot carries on with the next step, waiting for three minutes while the popcorn cooks. Anyone having encountered popcorn before can already tell what is wrong with this scene. A failure handling routine to detect and regrab a dropped lid would address only a symptom, not the root of the problem, *which is that the robot cannot leverage knowledge about the causal structure of the world into mechanisms to detect and repair failures as such failures arise*.

Our paper aims at addressing one of the parts needed to computationally implement such an ability: the need to represent practical world knowledge such that it is *transferable between situations*. To this end, one needs formal theories of functional relations and causal laws at a level that abstracts away from 'irrelevant' particulars, i.e. some qualitative descriptions of categories of scenes and causal laws. However, knowledge must also be grounded in procedures that can instantiate or recognize the instantiation of some general, qualitative pattern into a particular scene or arrangement of objects.

To bridge between these levels of abstraction, we draw on a formalisation of the notion of *image schemas* developed within cognitive linguistics (e.g., Johnson, Talmy [1,2]) and ontological formalizations pursued by Hedblom and others [3]. Image schemas are a plausible inventory of preconceptual building blocks of cognition, and can be used to describe functional relations between objects, i.e. relations whose applicability depends on the ability of objects to support certain behaviors, such as one object 'containing' another [4,5]. Such relations offer a useful level of description for characterizing expected or required participant behaviors and we formalize them by augmenting our ontological accounts with logical theories qualitatively describing spatial arrangements, expectations on behavior, as well as counterfactual expectations should required participants be absent. Formalizing appropriate levels of abstraction between generalized schemas and actual characterisation of physical locations and movements in the world (or in a simulation) allows us to define and address some core competency questions that an agent must engage with when acting in a changeable world: "what would happen if?", "why did something happen?", and "how can some state of affairs be brought about/avoided?" Being able to provide appropriate responses to such questions is a significant indication of an agent's understanding of the world in which it finds itself.

2. Background: Image Schemas

An image schema is defined by Johnson [1, p. xiv] as: "a recurring dynamic pattern of our perceptual interactions and motor programs that gives coherence and structure to our experience". Image schemas have attracted much attention in cognitive linguistics, and are thought to be involved in mental processes including metaphor construction and concept invention. Several authors have proposed lists of image schemas, with significant overlap among them; it thus seems likely that the number of image schemas in a complete inventory should not be very large. Nevertheless, until recently, formalizations of image schemas as theories amenable to computational implementation have been scarce.

This has now been addressed by Hedblom, Kutz, Neuhaus [3], and Hedblom, Kutz, Mossakowski and Neuhaus [6], who propose a first-order logic axiomatization of image schemas. This formalization has been used to explore both concept blending [7] and concept invention [8]. There remains, however, considerable work to be done towards developing a fully satisfying formal treatment. Firstly, we believe there are several requirements that current formalizations do not meet. At least some form of non-monotonic reasoning is needed: e.g., a table 'supports' a cup in a different way than a rack would. This raises questions concerning how best to distribute the various sources of knowledge that would enable flexible descriptions. Tractability must also be a prime concern. The difficulty of finding a formalism that can both answer commonsense reasoning problems and be well-behaved computationally is discussed by Davis [9].

Secondly, a purely logical model of image schemas appears to miss crucial properties that led originally to their very proposal. As defined by Johnson, image schemas capture patterns of *embodied* experience but, without proper grounding, logical models remain empty symbols. Geometric or physical concerns must enter into descriptions of image schemas but are awkward to capture using first-order logical machinery. E.g., in the well-known egg cracking problem (Morgenstern [10]), despite relatively complex axiomatization, only a handful of material resilience levels are described. Simply put, a purely logical approach uses the wrong tools for the job precisely because the critical contributions of embodiment are not available.

Simulation may then be a more appropriate tool for capturing intuitions about image schemas, but exhibits its own problems when used for commonsense reasoning; surveys addressing this topic are given by Davis and Marcus [11,12,13]. Simulation alone cannot answer basic questions such as what would be relevant to simulate. Reasoning is required, which we consider best provided by a logical component of a schematic description of the world *in addition* to any treatment in terms of simulations.

As a result, we propose the following requirements for a theory of image schemas; these requirements are meant to cover logical, as well as grounding, aspects of the theory:

- Non-monotonicity: allow defaults and exceptions when describing how an image schema would be instantiated in an arrangement of objects
- Well-foundedness: theories for schemas expressed in terms of simpler schemas, according to some principles of decomposability
- Tractability: crucial for agents acting in the physical world; implies the need for some sort of approximation or compromise in inferential power
- Correspondence to generative models: an image schema must be associated with procedures by which an agent can create instances of the schema
- Correspondence to perception procedures by which instantiations of image schemas can be recognized in an arrangement of objects.

We now detail our approach to formalizing image schema theories in a manner that is in line with these requirements and illustrate how we can then use this formalization to provide deepened understandings of the consequences of physical situations.

3. A Multi-stratal Ontological Treatment of Functional Relations

In this section we provide an overview of a pipeline for converting qualitative, underspecified descriptions of object arrangements into fully quantitatively specified scenes that

are amenable to simulation (section 3.1), and then converting the quantitative simulation results back into qualitative descriptions of behavior (section 3.2). We argue that such a pipeline establishes a powerful tool for constructing theories of functional relations using Containment and Support as examples; a more complete inventory of the functional relations to be covered is specified in the GUM-Space ontology [5]. A purely logical theory might suffice for 'typical' combinations of objects, but a one-size fits all formalized approach is untenable when faced with the extreme variation and contingency of the real world [14]. Such an account would not allow us to interrogate whether, e.g., containment relations can actually hold between objects of widely varying shapes and sizes.

Simulation offers an additional reasoning mechanism well-suited for such geometric or physical aspects [15]. Although we take to heart the arguments from Davis and Marcus [12] against simulation, we find their perspective overstated. Human beings make inferences about functional relations even in the presence of great uncertainty. People expect their clothes to be in their luggage after a plane trip, even if they do not know how the luggage moved and cannot simulate the clothes inside. It is also true that simulators depend on having a physical model and that model may be inaccurate, unstable, or fail to cover interesting physics (although in such technical regards, simulators are only getting better). But embodied understanding of functional relations is not an either-or between purely logical approaches and simulation; both have their place. An inference rule such as "things in locked containers tend to stay there" would, for example, justify a traveller's conclusion that despite its unknown trajectory, the luggage still contains the clothes.

Of course, such a rule has many exceptions – e.g., if the container has holes bigger than the contents. Attempting to formalize the entire complex of possible configurations and their consequences as abstract rules is consequently both infeasible and unlikely to cover new situations. In contrast, simulators encode knowledge about the physical world in a very compact form. As an alternative reconciliation of the either-or case, therefore, Bateman [14] discusses the need for allowing flexible selections of formalizations that more appropriately and systematically distribute explanatory work across hybrid formalizations. This would allow logical theories to be used to specify simulated 'introspection' concerning physical arrangements, whose results may then be interpreted back into propositions that can be reasoned with at the symbolic level. We detail the arrangement and interpretation of such mental simulation experiments in the next two subsections.

3.1. "Scene generation": From Qualitative Description to Fully Specified Arrangements of Objects

We approach the general simulation specification task using image schemas, formalizing these across several levels of abstraction. This allows us to ontologically characterize image schemas not just in terms of interdependencies of logical theories sharing the same formalism [3], but also in terms of the nature of the formalisms needed at each particular level of abstraction. We then work towards full specifications, by which we mean a quantitative description in which shapes, coordinates, and physical properties of the objects involved in a scene are given. The input for this process of refinement is a qualitative description expressed in terms of functional or spatial relations between objects; this is of necessity (and also usefully) comparatively underspecified; there are always several ways to instantiate a qualitative description. Our process of refinement then operates by relating information at each of the following levels of abstraction maintained.

Functional relations are spatial relations between entities that also constrain behavior. An inventory of functional relations is given by subconcepts of *FunctionalSpatialModality* from GUM-Space [5]. New to our account here is an explicit formalization of the Expectations a functional relation gives rise to, and counterfactual Expectations about what would happen if some participant in the relation would be removed. We draw a distinction between expectation in a colloquial sense, and Expectation as a qualitative description of behavior conditional on an entity, as used in our formalization – that is, a colloquial expectation would be that popcorn is contained in the pot it cooks in; this comes from cultural norms concerning how a well-performed cooking task unfolds. In contrast, the formal Expectation that the containee remains in a container is part of our formalization of the Containment relation, and is inferred as soon as a Containment relation is asserted without requiring additional consultation of norms, tasks or contexts.

A theory template for a functional relation is a set of defeasible Horn clauses where the terms are predicates parameterized by variables; a variable may appear as an argument for several predicates. To produce a theory for a functional relation holding between particular objects, all variables must be consistently replaced by identifiers referring to objects in some environment or entities from an ontology of spatial relations [5], resulting in a propositional defeasible logic theory. We select defeasible logic to account for exceptional ways in which a functional relation may be brought about, and here we will only consider inferences on the propositional theories resulting from instantiating templates. As a consequence, our system can reason about the consequences of statements such as "the popcorn is in the pot" but, because of the exclusion of logical quantification, does not consider statements such as "there exists something which contains the popcorn". This limitation is imposed to enable a clear separation of concerns between our system and more complex reasoners it may form a part of: our hybrid reasoning is a way to check to what extent a collection of propositions describing relations between specific objects is physically feasible, and to extract information, on physical grounds, about which other objects contribute to a relation, as described in our competency questions section; we note here that inference for propositional defeasible logic lies in P-time [16]. A task planner would be interested additionally in existentially quantified statements, e.g. whether there is some set of objects which can be arranged to obey a functional specification, and may then use our system to check candidate object sets.

A fragment of an example theory template is shown for Support in Listing 1, where $\Rightarrow$ denotes defeasible implication. The various predicates appearing on the righthand side correspond to lower layers of schemas defined in subsequent paragraphs. Capital single letters are variables that must be replaced when producing an instance of a theory, and "constants", i.e. parametrizations of predicates by entities from an ontology valid for all instances, are given in quotation marks. A Default Expectation describes what should happen when all participants are allowed to physically interact. A counterfactual Expectation describes what should happen if one of the participants does not physically influence others. The descriptions of observed behavior, such as SpecificDirectionalDown, will be presented in section 3.2.

Listing 1: Fragment of the theory template for Support

```
Support(X,Y) ⇒ Location(X,Y,'on')
Support(X,Y) ⇒ Expectation(Default(),SpecificDirectionalStayLevel(X,Y))
Support(X,Y) ⇒ Expectation(Disabled(Y),SpecificDirectionalDown(X,Y))
Support(X,Y) ⇒ ¬Support(Y,X)
```

Spatial relations are schematic relations that constrain the placement of objects in terms of geometric primitive relations, such as alignments, between their primitive features. Theories for spatial relations are also instantiated from templates, and a theory of a spatial relation holding between a collection of objects is a propositional defeasible logic theory. An example theory template for Locations with spatial modality 'on' (cf. [5]) is given in Listing 2.

Listing 2: Fragment of the theory template for Locations with spatial modality 'on'

```
Location(X,Y,'on') ⇒
    SurfaceContainment(
        ObjectRelativeBottomSurface(X),
        WorldRelativeTopSurface(Y))
Location(X,Y,'on') ⇒
    AxisAlignment(
        ObjectRelativeUpright(X),
        WorldRelativeUpright(Y))
Location(X,Y,'on') ⇒
    ¬Location(Y,X,'on')
```

Instances of theories for spatial relations operating at a lower level of abstraction often appear because a functional relation implies a spatial relation; e.g., the theory for $Support(cup, table)$ would imply $Location(cup, table, \text{'on'})$. We require that the parameters that can be accessed to create an instance for a spatial relation theory are constrained by the entities mentioned at the more abstract level of the functional relation theory, and the spatial relation must not depend on entities not mentioned at the more abstract level – that is, the theory for $Location(cup, table, \text{'on'})$ must not reference some other object apart from the cup and table. We impose this limit because otherwise we would effectively have existential quantification, which we wish to avoid because of the separation of concerns mentioned above, and to avoid combinatorial explosion.

Geometric primitive relations describe constraints on how geometric parts of objects may be arranged. They are not formalized as logical theories but rather implemented as numeric procedures to generate and filter a set of candidate placements using a probability distribution on spatial configurations. This approach is standard in robotics for representing regions; detailed presentations may be found, for example, in work describing the Cognitive Robotics Abstract Machine [17,18] or Action Related Places [19]. One feature of this approach is the ability to combine several constraints on object relative placement under a uniform representation – thus, probability distributions corresponding to different constraints, e.g., AxisAlignment and SurfaceContainment, can be combined into a single distribution, corresponding to the conjunction of constraints.

As they occupy lower abstraction levels than spatial relations, geometric primitive relations are typically invoked because they are implied by the theory of some spatial relation; e.g., $Location(cup, table, \text{'on'})$ implies

$$AxisAlignment(ObjectRelativeUpright(X), WorldRelativeUpright(Y))$$

This means that the entities participating in a geometric primitive relation must be a subset of the entities participating in the invoking spatial relation or their parts. The parts are obtained by invoking the next lower level of abstraction of the geometric primitives themselves, which are described next.

Geometric primitives are specified in terms of procedures to convert a shape description of an object into a representation of one of its parts or features. Examples include the centroid of an object, its outer surface, or its length axis. Primitives may be object- or world-relative, depending on which coordinate system they use. They can be determined by geometry (e.g. a PCA analysis can identify a longest axis) or asserted by convention (e.g. human objects are often designed with a particular direction intended as upright).

3.2. "Behavior interpretation": from quantitative simulation results back to qualitative descriptions

The behaviors we model currently as illustrative examples are represented by primitive movements, which are movements that describe the motion of some trajector object relative to another object, the relatum; these are covered by relational spatial modalities within the GUM-Space ontology [5]. In particular, the primitive movements we have considered so far are the two *GeneralDirectional* modalities 'Nearing' and 'Approaching', *SpecificDirectional* movements constrained to the vertical direction, and two further movement descriptions defined for modeling convenience: *RelativeMovement* and *RelativeStillness*.

We formalize primitive movements in terms of recognition procedures that take trajectory data as input – i.e., the relative pose of the trajector to the relatum at different time steps – and compute a cumulative cost over the duration of the input trajectory. The cost measures to what extent the actually observed trajectory deviates from the specification. This is not the same as deviating from some ideal trajectory, however. There is, for example, no ideal trajectory for RelativeApproach; instead, displacements that move the trajector away from the relatum are counted towards increasing the trajectory cost. If and only if the cost exceeds a threshold value is the observed trajectory deemed not to respect the primitive movement. The threshold is currently based on a fraction of the sum of the lengths – i.e., the longest axes – of the participating objects. This fraction may differ for different primitive movements, but currently we set it to a tenth of the sum of lengths of the relatum and trajector. This might be finetuned by a number of methods.

In the section following we proceed to the competency questions relevant for the new levels of formalization introduced and show how they can be computationally implemented within our system, building on the levels of representation defined.

4. Competency questions enabled by a multi-layered schematic approach to physics reasoning

To begin, we summarize our competency questions thus: "what would happen if?", "why did something happen?", and "how can some state of affairs be brought about/avoided?". These questions seem very natural, but they hide several sources of complexity.

One important set of concerns involves just how far into the future do we push a "what if?" question and how far into the past do we push a "why?" For the purposes of formalization, we must be explicit about our horizons. Why-questions also pose the problem of defining what counts as a cause. What-to-do questions are hard to solve in general, because planning is complex; it is more plausible that what humans do is learn routines which are appropriate to some class of situations, and to some degree adapt-

able. Finally, the level of abstraction at which these questions should be answered needs specification. For example, one can always analyse a cause in finer detail, assuming the data is there but, very often, we do not seem to care in our activities about the motion of many small component parts, and instead prefer high level descriptions. Fuller specifications of the competency questions at issue will now be listed, in each case showing how answering them is implemented.

"What if" questions. These questions are understood here as taking a qualitative description of an arrangement of objects, e.g. a pot contains cooking popcorn, and outputting a qualitative description of how the arrangement would naturally behave, e.g. the popcorn distances itself relative to the pot. Conversion from a qualitative description to fully specified arrangements of objects, including coordinates and initial velocities, proceeds down along the hierarchy of levels described in section 3.1 above.

As an example of a schematically described scene, let us consider:

Containment (container=pot, containee=popcorn).

The Containment schema is a functional relation, and so has a theory constraining both the spatial placement of the entities and expectations on their behavior as suggested above. The placement of the pot and popcorn is simply the spatial relation schema:

Location (relatum=pot, locatum=popcorn, spatial_modality='in')

which in turn further implies the following geometric primitive relation constraint:

VolumeContainment(big_volume=InteriorCavity(pot), small_volume=popcorn).

The VolumeContainment primitive relation guides sampling for positioning pot and popcorn such that the relevant geometric parts (an interior cavity in the case of the pot, and the popcorn itself) obey the volume containment constraint. A fully specified scene can then be simulated, and the trajectories of objects analyzed to check correspondence to the relevant movement schemas (cf. section 3.2).

An issue however appears at this second step: what schemas should be tested against the trajectory data? And since schemas have multiple participants, how does one know for which participants to do the test? This is related to the critique from Marcus and Davis [12] that a simulation, on its own, does not offer guidance about how it should be interpreted, or what objects or movements are significant. Indeed, the bare facts of an activity may have many interpretations, and people appear to select such interpretations based on an interpretive framework constructed from contextual expectations. Dropping a cup means one thing in an argument, and another when bringing a drink.

To model such interpretive frameworks, we take the approach that a "what if" question must itself specify the schemas to test for. In other words, what we understand as a "what if" question is to check which, if any, of some qualitative expectations on the behaviors of some objects will hold, assuming the objects are arranged to satisfy some qualitative spatial constraints. The expectations to check are then built into the qualitative description of the scene in terms of functional relations. In the pot and popcorn example in the context of cooking, the Containment relation has, among others, the expectation:

Expectation (condition=Default(), event=RelativeStillness(a=pot, b=popcorn)).

This guides the interpretation of the simulation by pinpointing just those objects and movements that are relevant, and establishes a criterion to judge said movement. In this case, some of the popcorn particles will actually escape the popcorn interior, thereby violating the RelativeStillness expectation of Containment.

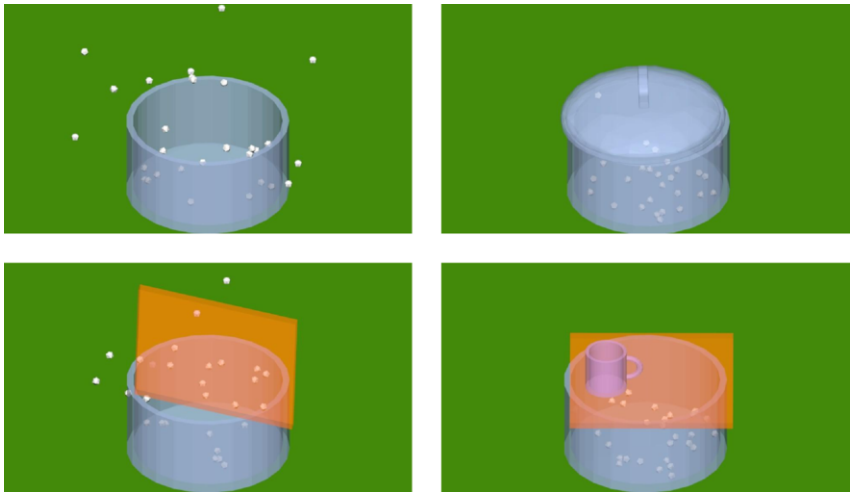

Figure 1. "What if" example scenarios: can various arrangements of objects contain popcorn?

Further examples of such "what if" questions follow for illustration. These are scenes involving pots, popcorn (which starts with some initial random velocity), lids, balsa boards (very lightweight), and cups. Screenshots of some frames from simulations relative to specific questions illustrative of the observed behaviors are shown in figure 1.

Question	Scene specification	Gloss of Result
"what if we tried to contain popcorn in a pot?"	Containment(pot, popcorn)	containment fail: popcorn flies out
"what if we tried to contain popcorn in a pot with a lid on top of it?"	Containment(pot, popcorn), Support(pot, lid)	ok
"what if we tried to contain popcorn in a pot with a light balsa board on top of it?"	Containment(pot, popcorn), Support(pot, balsa)	containment fail: popcorn flies out; support fail: balsa board does not stay level relative to pot
"what if we tried to contain popcorn in a pot with a light balsa board on top of it, and a cup on top of the balsa board?"	Containment(pot, popcorn), Support(pot, balsa), Support(balsa, cup)	ok

"Why" questions. These questions are understood here as attributing blame/credit to objects in a scene for the observed non/compliance of observed behavior to qualitative expectations placed on the scene by functional relations.

The approach we took to operationalize causality testing is interventionist [20,21]: an object can only be credited for a behavior if, by removing the object's influence from the scene, the behavior is no longer observed. Removing an object's influence means to stop it from interacting physically with other objects; we do not remove the object because its presence may be necessary for qualitative behavioral specifications, i.e., movement schemas, but we can readily prevent physical interactions. We refer to an object without physical influence as a phantom. Note that phantoms pose no problem for the simulator in terms of the physical consistency of the worlds created – they are simply ignored when performing updates of the physical state.

As an example of using phantom objects, let us consider two scenes, one in which we have a pot with cooking popcorn inside and covered by a lid, and another scene in which the pot is covered by a lid and contains a meditating pixie. In the former case, the popcorn particles have just popped and fly out. In the latter case, the pixie is simply content to levitate in place. The behavioral specification we are interested in for deriving potential causes is the prevention of a relative distancing between the pot and the popcorn or pixie. Obviously, both popcorn and pixie stay inside the pot if both the pot and lid are physically active, which by virtue of their properties as solid objects prevent objects from passing through. If, however, either pot or lid are phantoms, the popcorn will escape. In contrast, making the lid a phantom will still not result in the pixie leaving the interior region of the pot. Our system would then say that both the pot and the lid contribute towards keeping the popcorn near the pot, but the lid is not the cause of the pixie remaining in the pot.

Thus, in our framework, a "why" question must specify some schematic behaviors of interest, and a list of objects which may be responsible for those behaviors. Several scenes are simulated, one default scene in which all objects participate physically, and counterfactual scenes in which, in turn, one of the objects is turned into a phantom. Further illustrative examples of "why" questions follow. Having observed compliance (or not) to some functional specification of a scene in the previous section, we ask why the observed behaviors happened.

Question	Scene specification	Gloss of Result
"when a lid is on the pot, why does the popcorn stay inside the pot?"	Containment(pot, popcorn), Support(pot, lid)	both lid and pot are needed for containment
"a balsa board is on the pot; why does the balsa board fly off a pot with popcorn in it?"	Containment(pot, popcorn), Support(pot, balsa)	the popcorn is to blame, not the pot
"a balsa board is on the pot, and a cup is on the balsa board; why does popcorn stay in the pot?"	Containment(pot, popcorn), Support(pot, balsa), Support(balsa, cup)	pot, balsa board, and cup are all necessary for containment

Note that in the scenarios given, the containment relation is assured by interactions between objects which at the qualitative level are not explicitly asserted to contribute to the containment. This shows why adding a physical layer adds to the understanding of a situation beyond what a purely symbolic and qualitative approach is capable of. Some illustrative screenshots from simulations of 'counterfactual' scenarios are shown in figure 2. 'Phantom' objects are distinguished by black colors and higher transparency.

"What to do" questions. While such questions suggest planning, planning is expensive and in practice does not seem to be used often by humans. Instead, human beings learn simple rules of action, e.g. "to prevent popcorn from flying out of a pot, cover the pot". Answering a "what to do" question becomes a process of checking which such action rule might apply in a given situation, and what that might entail in terms of changes to an arrangement of objects. The focus of this competency question therefore is not on deliberation, but rather on the representational structure needed to know what are good responses to some class of situations. How these responses are acquired is a separate issue, for which we suspect pragmatic considerations are paramount. An agent might learn from instruction, or from observation, or even by simulated or actual experiments performed out of 'curiosity'.

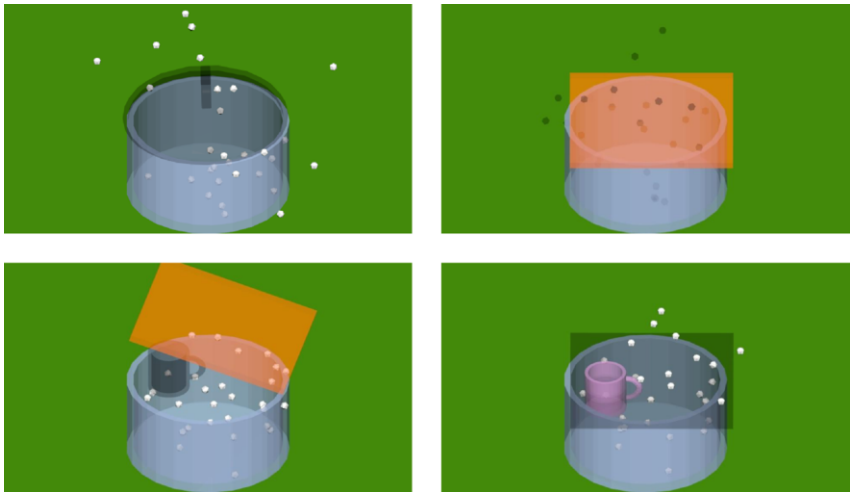

Figure 2. "Why" example scenarios: to check which objects contribute to an observed behavior, what happens when some objects have physical interactions disabled (marked by transparent black texture)?

To see how such a question could be answered, consider the example from the previous section of containing popcorn in a pot. Assume the following action rule: if an item of type pot fails to contain some other item X, then place another item Y on the pot. The $\mapsto$ symbol means a transformation from a scene qualitatively described by the left side to a scene qualitatively described by the right.

$$\neg\, Containment(X, pot) \mapsto Containment(X, pot) \wedge Support(Y, pot)$$

The rule antecedent, $\neg\, Containment(X, pot)$, can be asserted from prior knowledge or observed from real or simulated scenes. Deciding what object to add to a scene requires having a list of candidates to try out in simulation, e.g., small, manipulable items known to exist in the kitchen. A candidate is successful if the expectations of the functional relations are all met. Suppose then that possible candidates for Y are a cup, a plate, or a balsa board. Of these, only the plate achieves the intended result; the cup falls in the pot without stopping the popcorn; the balsa board is pushed off by the escaping popcorn.

Moreover, and as we have seen, combining items (e.g. the balsa board and the cup) can help the pot achieve containment too. Searching for such combinations of scene modifications might be done in an iterative deepening fashion, where modifications involving more objects are searched only if it is not possible to fix a scene with fewer items.

5. Related work

We refer to a recent survey by Davis and Marcus [22] for an overview of research into commonsense reasoning. By and large, commonsense reasoning has been pursued by way of attempting to construct either large repositories of facts – i.e., knowledge graphs – or ontologies [23], or rich logical theories often involving mixed formalisms to cover aspects such as time, geometry, topology [24]. Such logical approaches have been criticized, often by their own proponents [25,9], on the grounds of requiring intractable or undecidable formalisms, or, as in [14], on the grounds of over-commitments causing them

to need complex formalisms in the first place. Naturally, critiques of logical approaches to commonsense inference have also been made on the same grounds as critiques against purely symbolic approaches to AI in general [26,27].

A strong case can be made, however, that machine learning approaches fare no better. Although deep learning can construct agents that master a specific game, such agents do not have a conceptual understanding that would allow them to transfer their competence to even slightly modified versions of it (Kansky [28]). Kansky suggests "schema networks", a hybrid between a classical propositional logic and learning, as an approach to remedy this, but it is not yet clear if they would scale beyond the worlds of very simple Atari games. More complex sensorimotor concepts have been modelled as control policies or state estimation routines for partially observable Markov processes [29], but the way these policies are learned depends strongly on a curriculum, suggesting that knowledge of the concepts needs to be already formalized somewhere else, and in particular the dependencies of complex sensorimotor concepts on simpler ones must be explicitly known by whoever sets up the training protocol.

Spatial reasoning is also a very relevant area, comprising topics such as parthood [30] or region connectivity [31], quantitative reasoning about how a qualitative arrangement might be instantiated [32], and linguistically-motivated ontological modelling of spatial relations such as the GUM-Space ontology [5] or the theory of sense clusters [33]. Spatial calculi are also applied in Hedblom et al.'s latest characterizations of image schemas [34]. We have made considerable use throughout of previous work on image schemas, originally proposed by Johnson [1] as remarked above. We view our work as a continuation of the formalization attempts of Hedblom and others [3,6] in which we combine a logical formalism with geometric and simulation techniques to provide grounding for image schemas into generative models and recognition procedures. This then extends previous accounts towards interaction with embodied simulation as well.

An ontological characterization of causal/causal-like relations between individual occurrents has been provided by Galton [35]. On the practical side however, we have chosen to follow the treatment of causation offered by Pearl [20], i.e. an interventionist understanding of causation [21]: X can be a cause of Y if some change to X specifically results in a change to Y. As a result, in this work we analyze causal relations by tracking how particular modifications to a simulated scene alter the observed qualitative behavior.

6. Conclusions

An understanding of the physical world an agent is embodied in requires a hybrid formalism: one able to operate at a high level of abstraction, and hence generality, but also account for physical and geometric aspects of the world. A difficulty in creating such a hybrid is the tension between the need for underspecification when one aims for generally applicable knowledge, and the requirement for precisely quantified descriptions usable by tools for modelling physical interactions, such as simulation.

We resolve this tension by taking inspiration from image schemas, which are intended to be strongly related to embodied interactions as well as amenable to logical formalization. We propose a multi-layer formal approach, where each layer is characterized by a different level of abstraction and modelling task. The most abstract level is that of functional relations qualitatively describing expectations on object behavior and formal-

ized in terms of spatial relations and primitive movements. These are described, at lower levels of abstraction, via generative models to instantiate and recognize arrangements of objects that satisfy a qualitative description.

We use our formal theories of functional relations to answer questions about object arrangements, such as whether particular arrangements can enact a functional relation and why (not), and show that our approach allows a deeper understanding of such functional relations, including how background objects, not explicitly participating in the relation, contribute to it. We have also sketched how our approach could be used to describe response rules for an agent – what to do in particular situations in order to achieve some qualitative goal – but we leave further developments in this direction for future work.

Acknowledgments

The research reported in this paper was supported by the German Research Foundation (DFG), as part of the Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE - Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org/). The research was conducted primarily within subproject P01: 'Embodied Semantics for the Language of Action and Change'.

References

[1] Johnson M. The body in the mind: the bodily basis of meaning, imagination, and reason. University of Chicago Press Chicago; 1987.
[2] Talmy L. The fundamental system of spatial schemas in language. In: Hampe B, editor. From perception to meaning: image schemas in cognitive linguistics. Berlin: Mouton de Gruyter; 2006. p. 37–47.
[3] Hedblom MM, Kutz O, Neuhaus F. Choosing the Right Path: Image Schema Theory as a Foundation for Concept Invention. Journal of Artificial General Intelligence. 2015;6(1):21–54.
[4] Coventry KR, Garrod SC. Saying, seeing and acting. The psychological semantics of spatial prepositions. Essays in Cognitive Psychology series. Hove, UK: Psychology Press; 2004.
[5] Bateman JA, Hois J, Ross R, Tenbrink T. A Linguistic Ontology of Space for Natural Language Processing. Artif Intell. 2010 Sep;174(14):1027–1071.
[6] Hedblom MM, Kutz O, Mossakowski T, Neuhaus F. Between Contact and Support: Introducing a logic for image schemas and directed movement. In: 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017); 2017. p. 256–268.
[7] Hedblom MM, Kutz O, Neuhaus F. On the Cognitive and Logical Role of Image Schemas in Computational Conceptual Blending. In: Lieto A, Radicioni DP, Cruciani M, editors. AIC 2014 Artificial Intelligence and Cognition. vol. 1315 of CEUR Workshop Proceedings. CEUR-WS.org; 2014. p. 110–121.
[8] Hedblom MM, Kutz O, Neuhaus F. Image Schemas and Concept Invention. In: Confalonieri, R and Pease A, Schorlemmer M, Besold TR, Kutz O, Maclean E, Kaliakatsos-Papakostas M, editors. Concept Invention. Springer; 2018. p. 99–131.
[9] Davis E. Qualitative Spatial Reasoning in Interpreting Text and Narrative. Spatial Cognition & Computation. 2013;13(4):264–294. Available from: https://doi.org/10.1080/13875868.2013.824976.
[10] Morgenstern L. Beyond Toy Problems: A Logical Formalization of the Egg-Cracking Domain. In: Fourth International Symposium on Logical Formalizations of Commonsense Reasoning, http://www-formal.stanford.edu/leora/cs98/egg.a.ps; 1998. .
[11] Ludwin-Peery E, Bramley N, Davis E, Gureckis TM. Limits on the Use of Simulation in Physical Reasoning. In: Proceedings of the 41th Annual Meeting of the Cognitive Science Society, CogSci 2019: Creativity + Cognition + Computation, Montreal, Canada, July 24-27, 2019; 2019. p. 707–713.
[12] Davis E, Marcus G. The scope and limits of simulation in automated reasoning. Artificial Intelligence. 2016;233:60–72.

[13] Davis E, Marcus G. The Scope and Limits of Simulation in Cognitive Models. CoRR. 2015;abs/1506.04956. Available from: http://arxiv.org/abs/1506.04956.
[14] Bateman JA. Space, Language and Ontology: A Response to Davis. Spatial Cognition & Computation. 2013;13(4):295–314.
[15] Bateman J, Pomarlan M, Kazhoyan G. Embodied contextualization: Towards a multistratal ontological treatment. Journal of Applied Ontology. 2019;14(1):1–35.
[16] Maher MJ. Propositional Defeasible Logic Has Linear Complexity. Theory Pract Log Program. 2001 Nov;1(6):691–711. Available from: https://doi.org/10.1017/S1471068401001168.
[17] Beetz M, Mösenlechner L, Tenorth M. CRAM – A Cognitive Robot Abstract Machine for Everyday Manipulation in Human Environments. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Taipei, Taiwan; 2010. p. 1012–1017.
[18] Mösenlechner L, Beetz M. Fast Temporal Projection Using Accurate Physics-Based Geometric Reasoning. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). Karlsruhe, Germany; 2013. p. 1821–1827.
[19] Stulp F, Fedrizzi A, Mösenlechner L, Beetz M. Learning and Reasoning with Action-Related Places for Robust Mobile Manipulation. Journal of Artificial Intelligence Research (JAIR). 2012;43:1–42.
[20] Pearl J. Causality. 2nd ed. Cambridge University Press; 2009.
[21] Woodward J. Making Things Happen: A Theory of Causal Explanation. Oxford University Press; 2004.
[22] Davis E, Marcus G. Commonsense reasoning and commonsense knowledge in artificial intelligence. Commun ACM. 2015;58(9):92–103. Available from: https://doi.org/10.1145/2701413.
[23] Cycorp. OpenCyc 0.7.0; 2004. Available from: http://www.opencyc.org.
[24] Davis E, Marcus G, Frazier-Logue N. Commonsense reasoning about containers using radically incomplete information. Artificial Intelligence. 2017;248:46 – 84.
[25] Davis E. Naive Physics Perplex. AI Magazine. 1998 Dec;19(4):51.
[26] Churchland PM. On the nature of theories: a neurocomputational perspective; 1990. .
[27] Smith BC. The owl and the electric encyclopedia. Artificial Intelligence. 1991;47:251–288.
[28] Kansky K, Silver T, Mély DA, Eldawy M, Lázaro-Gredilla M, Lou X, et al. Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics. In: Precup D, Teh YW, editors. Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017. vol. 70 of Proceedings of Machine Learning Research. PMLR; 2017. p. 1809–1818.
[29] Hay N, Stark M, Schlegel A, Wendelken C, Park D, Purdy E, et al. Behavior Is Everything: Towards Representing Concepts with Sensorimotor Contingencies. In: McIlraith SA, Weinberger KQ, editors. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018. AAAI Press; 2018. p. 1861–1870.
[30] Casati R, Varzi AC. Parts and places: the structures of spatial representation. Cambridge, MA and London: MIT Press (Bradford Books); 1999.
[31] Randell DA, Cui Z, Cohn AG. A Spatial Logic Based on Regions and Connection. In: Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning. KR'92. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1992. p. 165–176.
[32] Bhatt M, Lee JH, Schultz C. CLP(QS): A Declarative Spatial Reasoning Framework. In: Egenhofer M, Giudice N, Moratz R, Worboys M, editors. Spatial Information Theory. Berlin, Heidelberg: Springer; 2011. p. 210–230.
[33] Bennett B, Cialone C. Corpus Guided Sense Cluster Analysis: a methodology for ontology development (with examples from the spatial domain). In: Garbacz P, Kutz O, editors. Formal Ontology in Information Systems - Proceedings of the Eighth International Conference, FOIS 2014, September, 22-25, 2014, Rio de Janeiro, Brazil. vol. 267 of Frontiers in Artificial Intelligence and Applications. IOS Press; 2014. p. 213–226. Available from: https://doi.org/10.3233/978-1-61499-438-1-213.
[34] Hedblom MM, Kutz O, Peñaloza R, Guizzardi G. Image Schema Combinations and Complex Events. KI – Künstliche Intelligenz. 2019;33:279–291.
[35] Galton A. States, Processes and Events, and the Ontology of Causal Relations. In: Formal Ontology in Information Systems - Proceedings of the Seventh International Conference, FOIS 2012, Gray, Austria, July 24-27, 2012; 2012. p. 279–292.