Formal Ontology in Information Systems
S. Borgo et al. (Eds.)

doi:10.3233/978-1-61499-910-2-141

The Mouse and the Ball

Towards a Cognitively-Based and Ontologically-Grounded Logic of Agency

Oliver KUTZ [a,1], Nicolas TROQUARD [a], Maria M. HEDBLOM [a], and Daniele PORELLO [a]

[a] *Conceptual and Cognitive Modelling Research Group (CORE), Faculty of Computer Science (KRDB), Free University of Bozen-Bolzano, Italy*

Abstract We discuss steps towards a formalisation of the principles of an agentive naïve proto-physics, designed to match a level of abstraction that reflects the pre-linguistic conceptualisations and elementary notions of agency, as they develop during early human cognitive development. To this end, we present an agentive extension of the multi-dimensional image schema logic ISL based on variants of STIT theory, thus replacing the temporal dimension of ISL with an action-agnostic theory of agency. To begin grasping the notion of 'animate agent', we apply the newly defined logic to model the image schematic notion of 'self movement' as a means to distinguish the agentive capabilities of a mouse from those of a ball. Finally, we outline the prospects for employing the theory in cognitive robotics.

Keywords. common sense reasoning, ontology of agency, spatio-temporal logic, image schemas, embodiment

1. Introduction and Context

Artefacts are reactive entities. Their behaviours produce the expected results given the appropriate setting and needed resources. A billiard ball would travel a straight line on the billiard cloth and fall into a pocket of the pool table, provided it is hit at the right point. Cognitive agents are proactive entities. A mouse set free on a pool table may also travel in a straight line into a pocket of the table. But the mouse does not need a hit, it can start its movement without external force. However, if it is hit and found running in a straight line it is more likely the result of trying to reach a safer place, not because it was pushed.

Toddlers and even pre-linguistic infants are able to comprehend the behavioural difference between the cue ball and that of the mouse [14]. They learn to predict that a ball pushed towards them will travel with no detour. They eventually fine-tune this understanding by learning that if not pushed hard enough the ball will stop before reaching them. It is, however, unlikely that they understand the physics behind it in terms of force, motion, momentum, and friction. It is suggested that infants make predictions about the world through an abstraction of their embodied experiences. These abstractions—image schemas—are sets of spatiotemporal object relations [9, 10]. Image schemas are con-

[1]Corresponding Author: oliver.kutz@unibz.it

ceptual primitives from which the principles of a naïve proto-physics can be derived: an intuitive understanding that humans form already in early infancy about objects in the physical world [20]. Image schemas target the pre-linguistic conceptual mind, differently from traditional efforts in qualitative reasoning or naive physics [6, 2]. Thus, when describing a simple situation such as a young child climbing and eventually sitting on a chair, it becomes essential to distinguish between the different affordances of objects (the chair gives SUPPORT) and the diverse capabilities of agents (the child can climb the chair–VERTICALITY). We here aim to motivate and study the principles of such naïve proto-physics and formalise corresponding pre-linguistic scenario conceptualisations and elementary notions of agency.

2. Image Schemas and Affordances

In simple terms, any object is an agent if it has the ability to perform actions. Developmental psychologists have long investigated how children come to form high level conceptual understanding such as the distinction between objects with agency and objects without [16]. During the first two years, children are thought to remain in a 'sensorimotor stage' where embodied experiences are thought to form the basis of future conceptual structure for higher level cognition such as language and analogical reasoning [19, 15]. Following this framework, image schemas were introduced as pre-linguistic conceptual structures generalised from the repeated exposure to particular spatiotemporal relationships between the self, objects and the environment [10, 9]. They capture relational concepts such as CONTAINMENT and LINK, dynamic ones such as SOURCE_PATH_GOAL and CYCLE, and force dynamic notions such as ATTRACTION and COMPULSION. It is believed that these patterns construe a conceptual skeleton for conceptual metaphors [11] and abstract concepts in mathematics [12]. The approach is related to Gibson's theory of affordances [5] in which the possibility for certain actions is described as the core of objects and agents, or as described in [3], where image schemas can be seen as 'bundles of affordances.' A tea cup is a tea cup because it affords the actions of containing and pouring in and out liquids from its form, basically capturing the image schema CONTAINMENT. A more agentive object, such as a cat, is a cat because it affords behaviours such as being pet, purring and has the ability for self movement.

These two cognitive theories, Image Schemas and Affordances, offer a first foundation upon which the distinctions between agents and inanimate objects, or tools, can be distinguished and identified. According to Mandler [15], the understanding of agency—both of the self and of others—and objects are developed from separate bases, namely one, from established action schemes, and two, from the analysis of observed data.

Some of the most fundamental differences between agents and inanimate objects that children learn as early as four months is to identify the difference between caused motion (CAUSED_MOVEMENT)[2] and self motion (SELF_MOVEMENT) [4]. Additionally, experiments with infants indicate that infants as early as five months have a concept of purposeful behaviour and devote more attention to actions with a perceived goal [23].

To separate inanimate objects and (animate) agents we start from the following observation: objects lacking agency are more readily perceived as tools upon which affordances can be executed, whereas agents have intentions of their own to perform actions.

[2]For formal work on the CAUSED_MOVEMENT image schema and related image-schematic notions see [7].

3. Agency: Ontology and Criteria

Our goal is the epistemic or cognitive classification of objects as behaving agent-like. For this reason, we endorse Dennett's criteria [21] to establish *agentivity* (or else) of an object relative to the behavioural information that a cognitive agent may collect. Among the different capabilities that an object can manifest in interacting with the environment, we concentrate on physical (spatial, temporal, and material) interactions leaving aside sophisticated, e.g. social, interactions. We leave aside also the simplest interaction in which an agent can be involved, that of sensing the environment, e.g. by listening or watching. In particular, we focus on how to ascribe intentionality and agency to objects that are moving in a certain environment.

Since we want to model the ascription of agency made by agents with possibly limited capacity, which may not have a fully developed theory of mind, actions, and intentionality at disposal, we do not want to commit to any substantive theory or ontology of action and intentionality. For this reason, we shall develop our modelling of actions within the tradition of the logic of agency [1]. As we shall see, this family of logics model in fact the result of an action on the environment, rather then committing to what an action is and how it can be characterised ontologically. *Logics of agency* have been developed precisely for investigating the principles of reasoning about actions, without committing to an *ontology of action*. The formulas that represent actions are in fact representing observable behaviours of an object in an environment that can be classified as agentive. This view fits the idea that agency is something that can be ascribed by looking at the behaviour of an object in an environment. Among the viable alternative logics of agency, we shall use Belnap's logic of "see to it that" STIT [1]. The reason for this choice is that Belnap relates the agentivity of an object to the non-deterministic nature of its behaviour. Non-determinism is given by the choices that an agentive object has, which prevents us from forecasting its future behaviour. Focusing on moving objects, for example, a mouse can be distinguished from a cue ball because the latter has no autonomous choice of changing its trajectory, which can happen only in the case of an external force that acts on it. This is evident from the cue ball's behaviour in the environment. By contrast, the mouse can change its trajectory in a manifestly autonomous way, with no external intervention. As we will see below, in STIT we can specify formally the distinction between how a cue ball may move and how a mouse may move.

4. A Logic for Directed Movement

The image schema logic ISL^M, first introduced in [7], is defined over the combined languages of RCC8 [17], QTC_{B1D} [22], cardinal direction (CD) [13], and linear temporal logic over the reals (RTL) [18], with 3D Euclidean space assumed for the spatial domain.

RCC8 is used for the spatial dimension and to talk about the topology of regions. For instance, RCC8 offers predicates like EC (externally connected), DC (disconnected), and NTPP (non tangential proper part). We also use cardinal directions: *Left*, *Right*, *FrontOf*, *Behind*, *Above* and *Below*.

The movement dimension is taken care of by atomic propositions of the form $O_1 \rightsquigarrow O_2$ (O_1 moves towards O_2's position), or $O_1 \hookleftarrow O_2$ (O_1 moves away from O_2's position).

For the temporal dimension, $\varphi U \psi$ reads as "φ holds, until ψ holds." Then $\mathbf{F}\varphi$ (at some time in the future, φ) is defined as $\top U \varphi$, and $\mathbf{G}\varphi$ (at all times in the future, φ) is defined as $\neg\mathbf{F}\neg\varphi$.

Example 1 *Here are two examples of well-formed sentences that can be written in the language of ISL^M defined in [7] (and might be considered true in specific scenarios).*

- $FrontOf(a,b) \wedge a \hookleftarrow b \longrightarrow \mathbf{F}\neg FrontOf(a,b)$ *'If a is in front of b and moves away from b, then at some point it will not be in front of b;*
- $NTPP(a,b) \wedge a \hookleftarrow b \longrightarrow \mathbf{F}DC(a,b)$ *'If a is inside b but moves away from it, it will eventually be outside b'.*

The logic ISL^M makes no distinctions between agents and inanimate objects. This means that objects are treated equally regardless of their movement patterns.[3]

For billiard balls this produces no direct problems. Static object relations can be described using a 'Two-object' family in which objects can be in contact with one another (see [7]) and the dynamic aspects of moving balls can be described using the presented logic together with specifications of the SOURCE_PATH_GOAL as captured in the PATH-following family (see [8]).

When a billiard ball gets hit by a pool cue it goes through a scenario that can be described using image schemas. First there is CONTACT between the two objects in which FORCE is transferred from the cue onto the ball. This captures the image schema of CAUSED_MOVEMENT, which can be defined as a complex and predictable form of SOURCE_PATH_GOAL. Basically, CAUSED_MOVEMENT is movement that is initiated through the impact with another object and that follows a clear (in most cases a straight) trajectory (see the eight-ball in Figure 1). The predictability of CAUSED_MOVEMENT is not per se defined by the destination or the goal. As the ball does not exhibit any active agency it is up to physical laws and the ball's momentum to determine how far it might reach. When a ball is pushed in a particular direction it will, with few exceptions, move in a straight line, or as determined by its physical environment.

In comparison, if a mouse is released on the billiard table, a completely different scenario unfolds. Even if the mouse is at rest when 'gently poked' by the cue, it is unlikely it will follow the predictable path demonstrated by a ball. Instead, it will most likely move 'randomly', potentially with the same goal in mind, the pocket, but with a less direct trajectory (see the mouse in Figure 1).

Following the reasoning that infants early on learn to distinguish between agents and inanimate objects [14], this dimension therefore deserves a place in a logic for image schemas, as we pursue next. The basic strategy is to re-use the static, non-temporal part of ISL^M, to describe spatial snapshot scenarios at a given timepoint.

[3]Moreover, it lacks a concept language to adequately distinguish between different kinds of objects and agents. Introducing this layer poses significant difficulties semantically and is left for future work.

5. Adding agency to $\mathbf{ISL}^M$

Logical modalities of *agency* aiming at modelling the results of an action have been widely studied in the literature, especially in practical philosophy and in multi-agent systems. We read the formula $\mathsf{Does}_a\varphi$ generically as "object a **sees to it** that φ." In fact, STIT theories [1] host a zoo of befitting variants of this modality. Furthermore, a modality of historical possibility $\Diamond$ allows us to express that something is possible $\Diamond\varphi$, or settled $\Box\varphi$. The ability to bring about a state of affairs is captured by $\Diamond\mathsf{Does}_a\varphi$.

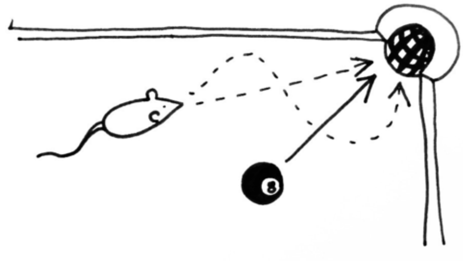

Figure 1. Potential movement pattern of a ball and a mouse.

An important aspect of STIT theory, contrary to dynamic logic, is that actions are not directly studied in STIT theory, i.e. the underlying ontology of STIT theory does not, in its standard formulation, assume concrete categories of actions or events [1]. We believe that this is a suitable choice for the basic cognitive modelling that we pursue in this paper.

We use the combined language of STIT and of ISL^M. STIT theories come with a rich semantics based on Ockhamist branching-time. The semantics for our combined language consists of the models of STIT theory (branching-time with agent choices) equipped with an interpretation function for the language of ISL^M. Statements about the future such as $\mathbf{F}EC(m,b)$ are evaluated w.r.t. a moment-history pair. The formula $\mathbf{F}EC(m,b) \wedge \Diamond\mathbf{G}\neg EC(m,b)$ thus means that at the current moment and history, there will eventually be contact between the mouse and the ball, but there is still a historical possibility for it not to happen.

A typical principle of folk physics is "what goes up must come down." If e stands for the Earth, and s is the sky, it can be formalised as:

$$Above(s,e) \wedge \mathbf{G}Above(s,e) \wedge x \rightsquigarrow s \rightarrow \Box\mathbf{F}x \rightsquigarrow e \ .$$

Such a statement, rather than being an axiom of the ISL^M logic, can be seen as an axiomatic constraint for the naive physics theory sketched in the introduction.

Semantics of the deliberative STIT logics The semantics of STIT is based on a branching time structure of the form $\langle W, < \rangle$, in which W is a *non-empty* set of moments, and $<$ is a *tree-like ordering* of these moments. A maximal set of linearly ordered moments from W is a *history*. A history being a set of moments, $w \in h$ indicates that the moment w is in the history h. We define $Hist$ as the set of all histories of a STIT structure. $H_w = \{h \mid h \in Hist, w \in h\}$ denotes the set of histories passing through the moment w. An *index* is a pair w/h, consisting of a moment w and a history h from H_w (i.e., a history and a moment in that history). In the following, $\mathsf{OAgt} = \{1,\ldots,n\}$ denotes a non-empty finite set of objects and Atm denotes a non-empty set of atomic propositions.

A *STIT model* is a tuple $\mathscr{M} = \langle W, <, Choice, v \rangle$, where: • $\langle W, < \rangle$ is a branching time structure, • $Choice : \mathsf{OAgt} \times W \to 2^{2^{Hist}}$ is a function mapping each object and each moment w into a partition of H_w, • v is valuation function $v : \mathsf{Atm} \rightarrow 2^{W \times Hist}$. The equivalence classes belonging to $Choice_a^w$ can be thought of as possible choices or actions available to object a at w. Given a history $h \in H_w$, $Choice_a^w(h)$ represents the particular choice from $Choice_a^w$ containing h, or in other words, the particular action performed by a at the index w/h. (Additional constraints are enforced, see [1].)

In the interest of replacing the RTL temporalisation with a STIT temporalisation, we assume a function map mapping every index to a model of these ground formulas.

A formula of the deliberative STIT logic is evaluated with respect to a model and an index. When p is a ground formula, $\mathscr{M}, w/h \models p$ iff $\mathsf{map}(m/h) \models p$. A proposition φ is settled ($\Box\varphi$) at an index m/h when it holds even if h is not the actual history that will unfold. $\mathscr{M}, w/h \models \Box\varphi$ iff $\forall h' \in H_w, \mathscr{M}, w/h' \models \varphi$. $\Diamond\varphi$ is defined in the usual way as $\neg\Box\neg\varphi$. The object a sees to it that φ when φ holds for every history in the set currently chosen by a. $\mathscr{M}, w/h \models \mathsf{Does}_a\varphi$ iff $\forall h' \in Choice_a^w(h), \mathscr{M}, w/h' \models \varphi$.

6. A Scenario in Cognitive Modelling: The Mouse vs. The Ball

Recall that we are interested in modelling spatio-temporal scenarios on the conceptual-cognitive level of a young infant, where complex notions of events, actions, and ontological categorisations of objects and their affordances are not yet fully developed. Indeed, a basic assumption of our research is that modelling this level of conceptualisation will be essential for artificial intelligent agents to bootstrap from 'simple observations' to 'event conceptualisation' and to be able to generalise across similar spatio-temporal situations.

In the following, we use b to designate a billiard ball, p to designate a pocket of the pool table, and m to designate a mouse.

The billiard ball is an object in OAgt and not an agent proper. In STIT, it is simply modelled in a way that $Choice_b^w = \{\{H_w\}\}$. That is, at every moment w, the billiard ball has one unique choice, which consists in selecting all the histories going through w. It cannot interfere with the course of nature. Formalised, for any proposition φ, it is globally true that $\mathsf{Does}_b\varphi \rightarrow \Box\varphi$: the billiard ball brings about something only if it is already settled. We can write $\Box\mathbf{G}(\mathsf{Does}_b\varphi \rightarrow \Box\varphi)$. A 'proper' agent, like the mouse, can interfere with the course of nature, possibly bringing about something that is not settled.

An object/agent a is *truly agentive* for a proposition φ when:

$$\Diamond\mathbf{F}(\Diamond\mathsf{Does}_a\mathbf{F}\varphi \wedge \Diamond\mathsf{Does}_a\mathbf{F}\neg\varphi) \ .$$

Agent a may never exercise its power to decide whether φ or $\neg\varphi$ is eventually true, but there is a history and a moment where it does. The image schema SELF_MOVEMENT is thus witnessed by a proposition φ and a moment where $\Diamond\mathsf{Does}_a\mathbf{F}\varphi \wedge \Diamond\mathsf{Does}_a\mathbf{F}\neg\varphi$ holds true.

In the model of Figure 2, the ball is hit at moment w_1. There are four possible outcomes. The histories h_3 and h_4 represent a situation where the cue ball is not hit towards the pocket, while in the histories h_1 and h_2 it is hit with enough force in the direction of the pocket. The mouse has one choice at moment w_1, and two choices at moments w_2 and w_3. The 'left' choices represent the mouse bumping into the ball. The 'right' choice represents the mouse leaving it alone. The ball has exactly one choice at every moment. In this model, the ball will eventually be in the pocket if the initial

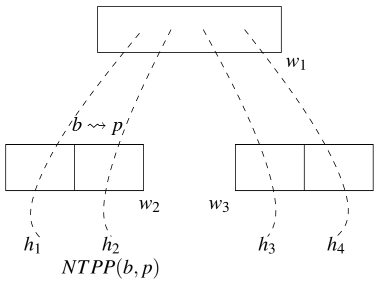

Figure 2. A billiard ball and a mouse.

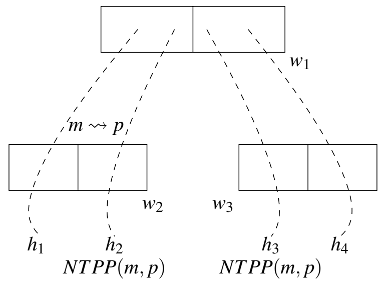

Figure 3. A self-moving mouse.

hit was straight and if the mouse chooses not to bump the ball. This course of action is represented by history h_2. If the billiard ball is moving towards the pocket, viz., at moment w_2, it cannot avoid to be in the pocket eventually:

$$b \rightsquigarrow p \rightarrow \neg\Diamond \mathsf{Does}_b \mathbf{F} \neg NTPP(b,p) \ .$$

Although, it cannot proactively avoid $NTPP(b,p)$, it is not a certain faith. There is still room for an interfering action of the mouse. The mouse can move into its trajectory. This is what happens at w_2 when the mouse chooses the choice $\{h_1\}$.

In the model of Figure 3, the mouse can decide at every moment whether to go towards the pocket ('left' choices), or not ('right' choices). After choosing $\{h_1,h_2\}$ at moment w_1, and finding itself in moment w_2, even if the mouse is moving towards the pocket, it can still make sure that it will never be in the pocket in the future.

$$m \rightsquigarrow p \rightarrow \Diamond \mathsf{Does}_m \mathbf{G} \neg NTPP(m,p) \ .$$

At moment w_3, the mouse can still change its mind and go towards the pocket anyway by taking the choice $\{h_3\}$. In fact, if the billiard ball and the mouse are moving towards each other, only the mouse can avoid contact:

$$b \rightsquigarrow m \wedge m \rightsquigarrow b \rightarrow \neg\Diamond \mathsf{Does}_b \mathbf{F} \neg EC(b,m) \wedge \Diamond \mathsf{Does}_m \mathbf{G} \neg EC(m,b)$$

The right hand side of the implication formalises the fact that the ball cannot bring about that eventually there will be contact, but the mouse can ensure that contact forever will not occur. This situation can be depicted by the very simple model depicted in Figure 4. At w_1, the mouse can choose to continue in the same direction, choice $\{h_2\}$, in which case the contact will occur. Otherwise, it can deviate by taking the choice $\{h_1\}$.

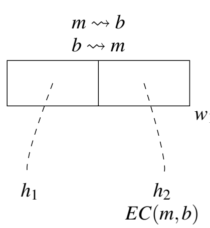

Figure 4.

References

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

[2] R. Casati and A. C. Varzi. 1997. *Spatial Entities*. Springer Netherlands, Dordrecht, 73–96.

[3] A. Galton. 2010. The Formalities of Affordance. In *Proc. of workshop Spatio-Temporal Dynamics*, M. Bhatt, H. Guesgen, and S. Hazarika (Eds.). 1–6.
[4] R. W. Gibbs, JR. and H. L. Colson. 1995. The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistcs* 6, 4 (1995), 347–378.
[5] J. J. Gibson. 1977. The theory of affordances. In *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, R. Shaw and J. Bransford (Eds.). NJ: Lawrence Erlbaum, Hillsdale, 67–82.
[6] P J Hayes. 1979. The naive physics manifesto. In *Expert systems in the micro-electronic age*, D Mitchie (Ed.). Edinburgh University Press.
[7] M. M. Hedblom, O. Kutz, T. Mossakowski, and F. Neuhaus. 2017. Between Contact and Support: Introducing a logic for image schemas and directed movement. In *Proc. of the 16th Int. Conference of the Italian Association for Artificial Intelligence (AI*IA 2017) (LNAI)*. Springer, Bari, Italy.
[8] M. M. Hedblom, O. Kutz, and F. Neuhaus. 2015. Choosing the Right Path: Image Schema Theory as a Foundation for Concept Invention. *Journal of Artificial General Intelligence* 6, 1 (2015), 22–54.
[9] M. L. Johnson. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago Press.
[10] G. Lakoff. 1987. *Women, Fire, and Dangerous Things*. Univ. of Chicago Press.
[11] G. Lakoff and M. Johnson. 1980. *Metaphors we live by*. The Univ. of Chicago Press.
[12] G. Lakoff and R. Núñez. 2000. *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics Into Being*. Basic Books, New York.
[13] G. Ligozat. 1998. Reasoning about Cardinal Directions. *J. Vis. Lang. Comput.* 9, 1 (1998), 23–44.
[14] J. M. Mandler. 1992. How to build a baby: II. Conceptual primitives. *Psychological review* 99, 4 (1992), 587–604.
[15] J. M. Mandler. 2004. *The Foundations of Mind: Origins of Conceptual Thought: Origins of Conceptual Though*. Oxford University Press, New York.
[16] J. Piaget. 1952. *The origins of intelligence in children*. NY: International University Press, New York. Translated by Margaret Cook.
[17] D. A. Randell, Z. Cui, and A. G. Cohn. 1992. A Spatial Logic based on Regions and Connection. In *Proc. 3rd Int. Conf. on knowledge representation and reasoning*.
[18] M. Reynolds. 2010. The complexity of temporal logic over the reals. *Annals of Pure and Applied Logic* 161, 8 (2010), 1063 – 1096.
[19] L. Shapiro. 2011. *Embodied Cognition*. Routledge, London and New York.
[20] B. Smith and R. Casati. 1994. Naive Physics: An Essay in Ontology. *Philosophical Psychology* 7, 2 (1994), 225–244.
[21] P. E. Vermaas, M. Carrara, S. Borgo, and P. Garbacz. 2013. The Design Stance and Its Artefacts. *Synthese* 190, 6 (2013), 1131–1152.
[22] N. Van De Weghe, A. G. Cohn, G. De Tré, and P. De Maeyer. 2006. A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and cybernetics* 35, 1 (2006), 97–119.
[23] A. L. Woodward. 1998. Infants selectively encode the goal object of an actor's reach. *Cognition* 69 (1998), 1–34. Issue 1.