Towards Reactive Robotics with a Pinch of Image-Schematic Reasoning

Mihai Pomarlan[1], Kaviya Dhanabalachandran[2] and Michael Beetz[2]

[1]*Applied Linguistics Department, University of Bremen, Germany*
[2]*Institute of Artificial Intelligence, University of Bremen, Germany*

Abstract
Today's robots do not possess a deep understanding of interactions between physical objects that is also available to their behavior generation modules and as such show brittle performance in realistic environments. While this suggests a robot would therefore need more knowledge, its decisions should not be too complicated to arrive at or else the robot risks losing track of what matters from its environment. Thus, we investigate a mix of reactive approaches to robotics and reasoning, and propose a simplified theory of typical changes between image schemas. We show how this theory could be integrated in a robot's perception-action loop, and describe some examples of using this theory to infer actions and perception queries for various stages of a pouring task. We are integrating this inference procedure into a simulated robot, but this integration is yet to be completed and as such, future work.

Keywords
rule based systems, cognitive robotics, commonsense reasoning, embodied cognition

1. Motivation

One of the hardest problems in AI-adjacent fields is to get an artificial agent to act in a competent and intelligent way in the physical world. While today's robots are testaments to the performance of control engineering, they lack a deeper understanding of the affordances present in their environment. Thus, we have industrial robot arms capable of precise control of the forces they apply [1], and robots that display great robustness in maintaining balance on two legs despite severe perturbations introduced by the kicks of onlookers [2, 3]. However, a robot – a general-purpose manipulator as opposed to a specialized device – that can autonomously cook an arbitrary recipe in a reasonable amount of time is beyond the current state of the art.

One cause for this state of affairs is that the problem of endowing robots with practical knowledge of affordances – how to detect them, how to use them – is yet to be solved by the robotics community. There is already research in discovering/learning affordances for robots [4, 5], and there are theoretical models of how affordances could be ontologically described [6]. However, to the best of our knowledge, such research has not completed a connection between theoretical descriptions of affordances and the behavior generation modules of a robot.

The Seventh Image Schema Day (ISD7), September 2, 2023, Rhodes, Greece
pomarlan@uni-bremen.de (M. Pomarlan); kaviya@uni-bremen.de (K. Dhanabalachandran); beetz@uni-bremen.de (M. Beetz)
0000-0002-1304-581X (M. Pomarlan); 0000-0002-0419-5242 (K. Dhanabalachandran); 0000–0002–7888–7444 (M. Beetz)

CEUR Workshop Proceedings (CEUR-WS.org)

The main question for a behavior generation module is what to do next; also important is to decide what next to pay attention to, out of the context of an activity[1]. In general, “what to do next” sounds like a planning problem, however planning is expensive and often unnecessary [8, 9]. Usually, in everyday activities it is obvious, at least for humans, what needs doing, and mistaken decisions can be rolled back at acceptable costs. Further, stopping to plan means paying less attention to how the environment changes for good or ill. Therefore, we will be interested in our line of work mostly on the reactive end of behavior generation techniques for robots. However, having a robot generate its behavior by hard-coded reactions that are too specific will make its performance brittle against what to us seem trivial changes.

Therefore we would like to endow our hypothetical reactive robot with some depth of knowledge about physical objects and how they interact. At the same time, this knowledge should not aspire to be a comprehensive description; rather, it is knowledge of what typically happens, and what typically to do. We can add to the previous requirements somewhat more formal ones. The robot’s theory of objects should offer the opportunity to abstract from situation details and use the abstractions to arrive at action/perception decisions; it should support inference chains of arbitrary length, if a situation demands it; but it should not require backtracking.

In this paper, we present our ongoing work in this direction. Specifically, we will present a simplified theory of image schemas whose aim is to describe how an agent would typically bring about the change of one situation, described image schematically, into another situation, also described image schematically. As examples, we illustrate our theory with various situations in which a pouring action is to be performed.

2. Related Work

How to represent knowledge for commonsense reasoning is an active topic of research. E.g., [10] introduces a general purpose knowledge management system able to incorporate symbolic and subsymbolic information. An in-depth formalization of a theory of containment in first-order logic is presented in [11]. In another work, a rule engine was used to infer the motions of an autonomous car [12] by first inferring possible behaviors then choosing one candidate in a conservative manner.

A promising building block for commonsense knowledge are image schemas. They are “dynamic analog representations of spatial relations and movements in space [...] derived from perceptual and motor processes” [13], and embodied patterns of sensori-motor experience [14]. As such, they offer a way to link situated, embodied activity with more abstract formal descriptions, as evidenced by the possibility to formalize image schemas in image schema logic (ISL) [15, 16]. Image schemas have been used to support hybrid, logical and simulation based reasoning, in [17], and provided a vocabulary in which to learn rules from simulations of pouring tasks in [18]. Previous work has suggested image schemas as a way to guide stacking actions [19] and robot action selection more generally [20]. Our paper here is a continuation and expansion of this latter line of work, in that we further develop the image schematic theory

---

[1] The definition of context from Turner is given as ”A context is any identifiable configuration of environmental, mission-related, and agent-related features that has predictive power for behavior. [7]”

to be used in action selection.

3. Setting: Controlling a (mostly) Reactive Robot

We first clarify our assumptions about who the primary user of our image schematic theory will be – i.e., what we mean when we say a robot is "(mostly) reactive".

A robot is a physical system able to "perceive"[2] and "act" in its environment in pursuit of goals. A reactive robot is one that can achieve this with no internal representation of its environment and with minimal complexity – it is "wired to do the right thing" [22, 8].

We believe however that in carving a niche for his reactive robotics research programme, Rodney Brooks overstated the case against the then established deliberative methods in robotics. We are not willing to abandon representations completely, therefore what we mean by "(mostly) reactive robot" is one which has the resources to construct, maintain, and use explicit descriptions of itself, its environment, and its goals at various levels of abstraction but which retains the other features of reactivity.

A (mostly) reactive robot only needs to consider the entities and relations that exist now, and what should exist as the completion of its goals. It will not search among alternative action sequences to achieve its goals, but instead only considers what the typical action is that would connect the present to the desired state of the world. It constantly monitors its environment and is able at any point to abort an action.

Finally, the perception of a robot, reactive or not, is sensitive to the robot's goals. This is because an environment can be described in any number of ways, each choosing different subsets of features to emphasize. Further, it is sometimes impossible to understand a sensory signal outside of a context that includes expectations and goals.

4. An Image Schematic Theory of Typical Change

We construct a – much simplified, compared to ISL – theory of image schemas in defeasible logic; we refer the reader to [23] for an overview of defeasible logic. Our logical theory is a collection of defeasible rules, where each rule has a set of predicates as antecedent and a predicate as consequent. The predicates are of arity at most two, and the rules cannot insert entities in the consequent that are not present in the antecedent.

The core entities in our theory are of course image schemas, understood as reified relations between participants that can be physical objects, features of such objects e.g. openings, or physical phenomena e.g. forces.

Participants in a schema play particular roles. E.g., a Near schema has a Locatum and a Relatum role. Note that while some schemas, such as Near, appear symmetric, we nonetheless prefer to have an asymmetry in the roles in that it is often not symmetric on which participant to act in order to bring about, or destroy the schema. If we say the cup, as locatum, is near the

---

[2] Psychological vocabulary such as "perceive", "goal" etc is to be interpreted metaphorically when applied to robots. We take an intentional stance [21] about sufficiently complex machines because it simplifies exposition and clarifies what human functions the various parts of said machines are intended to mimic – at the risk of suggesting said mimicry is successful.

table, the relatum, we don't just mean the symmetric fact of their spatial proximity but also that if we don't want these objects to stay near, it is the cup upon which we will act. Several relations can exist between schemas:

- *combine*: two image schemas coexist. They need not have participants in common. *combine* is a symmetric and transitive relation, and it is understood that together, the schemas in a combination describe a situation
- *follows*: an image schema exists after another in time. If s follows t, it cannot combine with t (but may combine with a schema of the same type and participants) and also follows every other schema that t combines with.
- *requires*: an image schema requires another one to hold before it can come into existence. If two image schemas combine, their requirements combine.
- *enables*: an image schema is followed by another one which it helps bring into existence.

A (mostly) reactive robot would use our theory by first creating a combination of image schemas that describe its goals. Already, the theory would be able to infer from this some other basic image schematic consequences – see the pouring example below – and what questions to ask of perception about the actual state of the world. Once available, perception results would also be described as a combination of image schemas, and a *follows* relation would be asserted between schemas in the goal description, and schemas in the actual state description. Further, for each schema s_k in the goal description, an individual d_k would be created and the relationship $requires(s_k, d_k)$ would be asserted. The theory is then able to identify the types and participants of image schemas d_k, as well as further questions to ask of perception. In a further loop, depending on whether schemas d_k are found to hold in the actual state, either *enables* relations are asserted or further requirements are introduced, this time for the schemas d_k themselves.

Thus, at every iteration of its perception-inference-action loop, the robot would start from its top-level goals and their enablers, and the results of the previous perception queries. It will then set up an inference problem for our theory with the goals and requirements as a combination of image schemas representing a goal situation, and the perception results and enablers as a representation of the actual situation. The result of inference is a new set of perception queries to ask for the next step, as well as what the current requirements are to achieve the goals, some of which map actions to undertake (other requirements may be that certain natural processes unfold, e.g. falling, where the robot should enable this unfolding through its actions).

Note that typically an image schema may have several requirements; however, it is often the case that these requirements should be fulfilled in sequence, hence it is enough to consider only one requirement at a time, and treat already fulfilled requirements as enablers. The fact that a schema is an enabler will result in perception queries being generated about it – we want to monitor that the parts of the world state that are helping us achieve a goal stay the way we want them to.

4.1. Example: Pouring

We provide the image schematic theory and some examples of its application in an openly accessible repository[3] which the reader is kindly invited to consult for more details.

The examples consist of a pouring task, where the goal is for “coffee” to be inside a “bowl” and outside a “cup”. Already from this combination of image schemas, our theory can also conclude that the cup should not be in the coffee. The examples then show the steps through which a robot will construct a tree of intermediary goals and perception queries.

- if the actual state is the coffee in the cup, then it must exit the cup, and it will do so through a waypoint that is the cup’s opening
- if the coffee must exit the cup, then the cup’s opening must not be blocked
- if the opening is not blocked, then typically it will be gravity that impels a fluid to exit a container
- if gravity is acting on an object, that object is falling
- to fall from a source through a waypoint, the source must be above the waypoint – from this, it is then inferred that the cup must be tipped in order to start pouring
- to fall into a destination through a waypoint, the waypoint must be above the destination – from this, it is then inferred that the bowl should be upright before we can start pouring

Further, every time an image schema describing placement or movement is identified as a goal, requirement, or enabler, appropriate perception questions are generated about its participants.

Note that the inferences in our example are not specific to the coffee pouring situation in our example. Rather, they are fairly general inferences about the operation of containers, the nature of falling, nearness and aboveness, and the typical choice of physical interaction to cause a particular kind of movement in some types of objects.

5. Conclusion and Future Works

In this paper, we have articulated the inference needs for a kind of robotic behavior generation that we called “(mostly) reactive” and how they could be fulfilled by a theory of typical change between image schemas. We have given an example of such a theory via several situations related to pouring, but the theory actually contains some more general knowledge about container behavior and some spatial, movement, and force schemas, in particular those related to verticality.

The theory itself is in development and we are adding more knowledge about the other image schemas from the classical lists. Also, we are in the process of integrating our inference system into the control architecture of a simulated robot that aims to perform a pouring task in a fairly realistic environment, though one in which perception is more powerful than real robot perception would be.

We do aim to eventually integrate our approach on a real robot, but we expect the problem of formulating perception queries, and of interpreting their results, to be significantly more

---

[3] https://github.com/mpomarlan/silkie and the theory and examples are in the examples/pouring_is folder of this repository.

difficult. On the other hand, one of the core assumptions of our approach – that perception should not produce some comprehensive ground truth but rather a good-enough for a specific purpose summary while being informed by top-down expectations – does have, in our opinion, the chance to ameliorate some of the difficulties of robot perception.

Acknowledgements

The research reported in this paper has been partially supported by the German Research Foundation DFG, as part of Collaborative Research Center (Sonderforschungsbereich) 1320 "EASE - Everyday Activity Science and Engineering", University of Bremen (http://www.ease-crc.org/).

References

[1] C. Ott, M. A. Roa, F. Schmidt, W. Friedl, J. Englsberger, R. Burger, A. Werner, A. Dietrich, D. Leidner, B. Henze, et al., Mechanisms and design of dlr humanoid robots, Humanoid Robotics: A Reference (2016) 1–26.

[2] E. Guizzo, By leaps and bounds: An exclusive look at how boston dynamics is redefining robot agility, IEEE Spectrum 56 (2019) 34–39.

[3] S. Liu, G. Lever, Z. Wang, J. Merel, S. A. Eslami, D. Hennes, W. M. Czarnecki, Y. Tassa, S. Omidshafiei, A. Abdolmaleki, et al., From motor control to team play in simulated humanoid football, Science Robotics 7 (2022) eabo0235.

[4] S. Rezapour Lakani, A. J. Rodríguez-Sánchez, J. Piater, Towards affordance detection for robot manipulation using affordance for parts and parts for affordance, Autonomous Robots 43 (2019) 1155–1172.

[5] D. Kraft, R. Detry, N. Pugeault, E. Başeski, J. Piater, N. Krüger, Learning objects and grasp affordances through autonomous exploration, in: Computer Vision Systems: 7th International Conference on Computer Vision Systems, ICVS 2009 Liège, Belgium, October 13-15, 2009. Proceedings 7, Springer, 2009, pp. 235–244.

[6] D. Beßler, R. Porzel, P. Mihai, M. Beetz, R. Malaka, J. Bateman, A formal model of affordances for flexible robotic task execution, in: Proc. of the 24th European Conference on Artificial Intelligence (ECAI), 2020. URL: https://ai.uni-bremen.de/papers/bessler20affordances.pdf.

[7] R. M. Turner, Context-mediated behavior for intelligent agents, International Journal of Human-Computer Studies 48 (1998) 307–330.

[8] R. A. Brooks, Planning is just a way of avoiding figuring out what to do next, Technical Report Working Paper 303, 1987.

[9] R. Brooks, A robust layered control system for a mobile robot, IEEE Journal on Robotics and Automation 2 (1986) 14–23. doi:PATH10.1109/JRA.1986.1087032.

[10] M. Beetz, D. Beßler, A. Haidu, M. Pomarlan, A. K. Bozcuoğlu, G. Bartels, Know rob 2.0—a 2nd generation knowledge processing framework for cognition-enabled robotic agents, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 512–519.

[11] E. Davis, G. Marcus, N. Frazier-Logue, Commonsense reasoning about containers using

radically incomplete information, Artif. Intell. 248 (2017) 46–84. URL: https://doi.org/10.1016/j.artint.2017.03.004. doi:Path10.1016/j.artint.2017.03.004.

[12] F. Bouchard, S. Sedwards, K. Czarnecki, A rule-based behaviour planner for autonomous driving, in: Rules and Reasoning: 6th International Joint Conference on Rules and Reasoning, RuleML+ RR 2022, Berlin, Germany, September 26–28, 2022, Proceedings, Springer, 2022, pp. 263–279.

[13] R. W. Gibbs, H. L. Colston, The cognitive psychological reality of image schemas and their transformations 6 (1995) 347–378. URL: https://doi.org/10.1515/cogl.1995.6.4.347. doi:Pathdoi:10.1515/cogl.1995.6.4.347.

[14] M. Johnson, The body in the mind: The bodily basis of meaning, imagination, and reason, The body in the mind: The bodily basis of meaning, imagination, and reason, University of Chicago Press, Chicago, IL, US, 1987. KerkoCite.ItemAlsoKnownAs: 2129771:M9VYBAD7.

[15] M. M. Hedblom, O. Kutz, T. Mossakowski, F. Neuhaus, Between contact and support: Introducing a logic for image schemas and directed movement, in: AIxIA 2017 Advances in Artificial Intelligence: XVIth International Conference of the Italian Association for Artificial Intelligence, Bari, Italy, November 14-17, 2017, Proceedings 16, Springer, 2017, pp. 256–268.

[16] M. M. Hedblom, Image schemas and concept invention: cognitive, logical, and linguistic investigations, Springer Nature, 2020.

[17] M. Pomarlan, J. A. Bateman, Embodied functional relations: A formal account combining abstract logical theory with grounding in simulation., in: FOIS, 2020, pp. 155–168.

[18] M. Pomarlan, M. M. Hedblom, R. Porzel, Panta rhei: Curiosity-driven exploration to learn the image-schematic affordances of pouring liquids., in: AICS, 2021, pp. 106–117.

[19] K. Dhanabalachandran, M. M. Hedblom, M. Beetz, A balancing act: Ordering algorithm and image-schematic action descriptors for stacking objects by household robots, in: The Eighth Joint Ontology Workshops (JOWO22), RobOntics: 3rd International Workshop on Ontologies for Autonomous Robotics Jönköping, Sweden, August 15-19, 2022, CEUR-WS, 2022.

[20] M. M. Hedblom, M. Pomarlan, R. Porzel, R. Malaka, M. Beetz, Dynamic action selection using image schema-based reasoning for robots., in: JOWO, 2021.

[21] D. C. Dennett, The Intentional Stance, The MIT Press, Cambridge, MA, 1987.

[22] R. A. Brooks, Intelligence without representation., Artif. Intell. 47 (1991) 139–159. URL: http://dblp.uni-trier.de/db/journals/ai/ai47.html#Brooks91.

[23] H. P. Lam, On the Derivability of Defeasible Logic, Phd, School of Information Technology and Electrical Engineering, The University of Queensland, 2012.