Back to the Future of Integrated Robot Systems

Mohan Sridharan

Institute of Perception, Action, and Behavior
School of Informatics
University of Edinburgh, UK
m.sridharan@ed.ac.uk

Abstract

Robots are increasingly being used in different application domains due to rapid advancements in hardware and computational methods. However, state of the art methods for many problems in robotics are based on deep networks and similar data-driven models. These methods and models are resource-hungry and opaque, and they are known to provide arbitrary decisions in previously unknown situations, whereas practical robot application domains require transparent, multi-step, multi-level decision-making and ad hoc collaboration under resource constraints and open world uncertainty. In this paper, I argue that for widespread use of robots, we need to revisit principles that can be traced back to the early pioneers of AI. We also need to make these principles the foundation of the architectures we develop for robots, with modern data-driven methods being one of many tools that build on this foundation. I then illustrate the potential benefits of this approach in the context of fundamental problems in robotics such as visual scene understanding, planning, changing-contact manipulation, and multiagent/human-agent collaboration.

1 Motivation and Claims

Robots are increasingly being deployed in application domains such as navigation, healthcare, and manufacturing. Although the development of high-fidelity hardware has aided this deployment, advancements in AI algorithms have revolutionized the field of robotics. In particular, methods and frameworks based on *end-to-end*, *data-driven*[1] deep networks and foundation models such as Large Language Models (LLMs) and Vision Language Models (VLMs) are considered state of the art for perception, reasoning, manipulation, and interaction problems in robotics (and AI) (Doshi et al. 2024; Huang et al. 2023; Schick et al. 2023; Surís, Menon, and Vondrick 2023; Zhang et al. 2024; Zhao, Lee, and Hsu 2023). There is a lot of hype and fear surrounding the development and use of such methods and models, with researchers claiming that these methods possess capabilities such as "planning", "commonsense reasoning", and "general intelligence". As a result, we are witnessing a rapid decline in the diversity of mathematical formulations being pursued to address open problems in robotics.

To motivate the need for a different and broader approach, let us first consider the key requirements characterizing integrated robot systems sensing and (inter)acting in the physical world. These requirements include the ability to:

- make multi-step, multi-level decisions based on multi-modal inputs such as vision, speech, and touch;
- operate under open world uncertainty, where optimal decisions are unknowable and probabilities often do not meaningfully model the uncertainty;
- operate under (often strict) constraints on resources such as computation, storage, and training examples;
- rapidly and incrementally augment and revise (as needed) existing models for various tasks such as perception, planning, and navigation; and
- support transparency in decision making, expressing these decisions in terms of human concepts such as beliefs and goals to promote understanding.

Let us next consider the well-known characteristics of modern deep network methods and models.

- They are excellent statistical predictors for well-defined tasks, but they may make the correct decisions or arbitrary ones in truly novel situations;
- Despite the development of architectures with different structures and properties, they are based on a narrow set of representations and update processes;
- They are resource-hungry, making substantial computation, data, storage and energy demands; and
- They are *batch learning* systems whose internal operation remains opaque; even when we are able to attribute decisions to specific nodes, we are often unable to ascribe meaning to this finding.

Even when such deep networks are used to develop *hybrid methods*, for example, the rich literature in neurosymbolic (NeSy) AI (Besold et al. 2022) and probabilistic NeSy AI (Smet et al. 2023), we end up with symbolic methods and/or probabilistic uncertainty models guiding the learning

[1]The use of these terms to characterize modern AI methods is actually incorrect; the design of end-to-end methods involves a considerable amount of "engineering" based on knowledge of task and domain. Also, there is an established history of AI algorithms learning from data. These terms are still used in this paper just to make better contact with existing literature.

of a deep network *backbone*. There is thus a fundamental mismatch between the requirements of integrated robot systems and the characteristics of the modern AI methods being developed for core problems in robotics. The associated representational choices and update processes limit expressivity, efficiency, transparency, and reproducibility. Furthermore, they contribute to the ongoing mad rush to set up large computing centers to collect and process all the data, potentially supported by many new power plants, leading to a substantial negative impact on sustainability.

2 Revisiting Key Principles

I argue that the mismatch described in the previous section can be addressed by revisiting some key principles that can be traced back to the early pioneers of AI but are not fully leveraged in modern robotics research. These pioneers were deeply inspired by (and contributed to) related disciplines such as Philosophy and Psychology, and much of their work in AI was inspired by a joint exploration of *natural intelligence*, i.e., cognition and control in humans and other biological systems. For example, the following observations are highly relevant to robotics research:

1. Human behavior is jointly determined by internal cognitive processes and the environment. We jointly explore the underlying perception, reasoning, control, and learning problems using *different representations and processes at different abstractions* (Sloman 2012; Turing 1952), automatically directing *attention* to relevant representations and processes as needed (Broadbent 1957; Triesman and Gelade 1980).
2. Unlike the "batch learning" and optimization approach currently prevalent in AI and other disciplines, humans acquire skills *incrementally*, *interactively*, and *compositionally* through *adaptive satisficing* under resource constraints and open world uncertainty; humans seek to make *rational* decisions instead of optimal ones (Gigerenzer 2021; Simon 1956).
3. Human skills, particularly our motor control skills, have evolved over a long time for some very hard engineering problems. It is difficult and computationally expensive to replicate these skills in robots (Minsky 1986; Moravec 1990). Also, just replicating our hardware, e.g., our arms and hands, is unlikely to be sufficient for the desired functional capability, e.g., dexterous robot manipulation.

These observations do not preclude the use of deep network or other similar methods in robotics. Instead, they advocate that we embed certain key principles in robot architectures, and consider deep networks as one of many different tools that can be included in the architecture as needed. Here, I highlight three such key principles.

1. **Refinement.** It can be viewed, in the context of robotics, as representing actions and change in the domain in the form of transition diagrams at different abstractions, with the fine(r)-granularity description being a *refinement* of the coarse(r)-granularity description. It is a fundamental concept that has appeared in research in robotics and computing over many decades. For example, in the field of software engineering and programming languages, there are approaches for type and model refinement, although they do not consider theories of actions and change (Freeman and Pfenning 1991; Lovas and Pfenning 2010; Mellies and Zeilberger 2015). To adapt this principle to robotics, the key idea is to establish a formal relationship between the descriptions at different abstractions that each support different representations and processes to update these representations. The relevant representations (and processes) can then be chosen automatically for any given task and domain depending on the information sources and resources available, by drawing on the other two principles described below. Note that even a limited exploration of this idea of selective attention (Broadbent 1957) has led to impressive results with deep networks (Doshi et al. 2024). Furthermore, the support for different representations enables the robot to incrementally acquire domain knowledge from different sources, and to interactively provide on-demand descriptions of its decisions in different ways that make contact with human concepts such as goals and beliefs.
2. **Ecological Rationality (ER).** It builds on Herb Simon's definition of *Bounded Rationality* (Simon 1955, 1956) and the related rational theory of heuristics (Gigerenzer 2020). Unlike the focus on optimal search in many disciplines (e.g., finance, computing) in the presence of *risk* over a set of known scenarios, ER studies decision making under *open world uncertainty*, i.e., when the space of possible scenarios is not known in advance. It characterises the behavior of a human or an AI system as a joint function of the internal cognitive processes and the environment, using *adaptive satisficing* to make rational decisions instead of optimal ones. Also, unlike the use of heuristics as a "hack" or to explain biases (e.g., in the *heuristics and biases* program in Psychology), ER considers heuristics as a strategy to ignore part of the information in order to make decisions more quickly, frugally, and accurately than complex methods (Gigerenzer and Gaissmaier 2011). Unlike modern AI research that is largely *prescriptive* (focusing on what should be done), it is both *descriptive* (describing what people or agents do) and prescriptive. It advocates an adaptive toolbox of classes of simple *decision heuristics* such as tallying, sequential search, and fast and frugal (FF) trees, and an algorithmic approach involving out-of-sample and out-of-population testing to identify heuristics that match domain characteristics. Such decision heuristics are well-suited to make decisions under open world uncertainty, where optimal decisions are unknowable and probabilities are not a good model of the uncertainty. Their design also automatically supports process-level explanations of the decisions made.
3. **Interactive learning.** It is a term used to refer jointly to different types of learning such as supervised (or unsupervised) learning and learning from reinforcement (Laird et al. 2017). The difference lies in how this learning is triggered and achieved. Modern AI systems focus on learning a single model or policy that deter-

mines decisions across different categories, situations, platforms, and/or domains. Such an approach is considered to be essential for *generalization* without realizing that there is a mismatch between the underlying design choices and the desired functional capabilities. For example, the learned model or policy is hard to understand, explain, or revise in a meaningful manner. These approaches are well-suited for tasks or domains in which the range of options or situations to be considered are known a priori and sufficient resources are available; they are not really suitable for decision making *in the wild*, i.e., under open-world uncertainty (Katsikopoulos et al. 2021a). Interactive learning, on the other hand, focuses on learning as needed to adapt to any given domain and set of tasks. Also, it advocates reasoning with any prior domain knowledge to inform and constrain the learning. Such an approach, not surprisingly, leads to simpler models that are amenable to incremental and rapid revisions, even in previously unknown situations, particularly when used in conjunction with the principles described above.

3 Architectural Examples

This section provides examples of embedding the principles outlined above in robot architectures to address problems in reasoning, control, collaboration, and learning.

Refinement for knowledge representation and reasoning. One example of refinement of agents' action theories used situation calculus to describe the theories, and assumed the existence of a bisimulation relation between the theories for a given refinement mapping (Banihashemi, Giacomo, and Lesperance 2017, 2018). Although assuming the existence of a bisimulation relation often has a negative impact on expressivity and computational efficiency in robotics domains, this work provides a good example of the transfer of information and control between two abstractions. There has also has been related work on combining discrete and continuous planning at different resolutions for task and motion planning (TAMP) in robotics (Garrett et al. 2021). This can involve using classical planners based on first-order propositional logic for planning discrete abstract movement actions, implementing each abstract action using continuous planners (Srivastava et al. 2013). This can also involve learning feature-based state and action abstractions towards generalized TAMP for continuous control tasks (Curtis et al. 2022).

Key limitations of the existing work are that they do not fully: (a) support the bidirectional flow of relevant information between the different abstractions; (b) handle uncertainty, particularly the effect of non-stationarity (of the domain) and future state uncertainty on the associated models; and (c) address the discontinuities in the interaction dynamics, i.e., the sudden changes in forces and the resultant acceleration experienced by the robot when it makes or breaks contact with objects and surfaces (Garrett et al. 2021).

My work has explored the hypothesis that the above-mentioned limitations are the result of not jointly considering the reasoning and learning problems, and not leveraging all the principles outlined above in the design of the corresponding architecture. For example, we developed a refinement-based architecture that supports different representations (logics, probabilities) and processes (non-monotonic logical reasoning, probabilistic sequential decision making) for reasoning with the transition diagrams at different abstractions (Sridharan et al. 2019). In particular, we have focused on making rational decisions; relaxing the need for performance guarantees; embedding cognitive theories of intention (Gomez, Sridharan, and Riley 2021), affordance (Langley, Sridharan, and Meadows 2018; Sridharan, Meadows, and Gomez 2017), and explainable agency (Langley et al. 2017; Sridharan and Meadows 2019; Sridharan 2024); and on automatically determining the part of the finer-resolution diagram relevant to implement any given coarse-resolution transition. We demonstrated experimentally that the resultant architecture performs better than state of the art knowledge-based or data-driven systems.

Decision heuristics for manipulation and collaboration. ER and decision heuristics have been used to achieve good performance on prediction problems in application domains such as finance, healthcare, and law (Brighton and Gigerenzer 2012; Durbach et al. 2020; Gigerenzer 2016; Katsikopoulos et al. 2021b). There is hardly any use of these methods in robot architectures, except in some related work in the cognitive systems community (Langley and Katz 2022). This lack of uptake is potentially due to their inherent simplicity, which is a strength but makes researchers doubt their suitability for practical problems. Also, unlike modern data-driven AI methods, the successes of decision heuristics do not receive the attention they deserve.

I present two examples of my work to demonstrate the power of decision heuristics. The first one focuses on collaboration between agents without prior coordination, i.e., *ad hoc teamwork* (AHT) (Mirsky et al. 2022). Methods considered state of the art for AHT use a large labeled dataset of prior observations to model the behavior of other agent types and to determine the ad hoc (AI) agent's behavior (Barrett et al. 2017; Rahman et al. 2021; Santos et al. 2021). As stated earlier, such methods do not support rapid incremental revisions or transparency, and the necessary resources (e.g., training examples, computation) are often not available in practical domains. In a departure from these methods, we adapted our refinement-based architecture to pose AHT as a joint reasoning and learning problem. This architecture enabled an ad hoc agent to choose its actions based on non-monotonic logical reasoning with prior domain knowledge (of action theories in different abstractions) and models learned and revised rapidly to predict the behavior of other agents. These predictive models were based on an ensemble of FF trees and used orders of magnitude fewer examples (e.g., 5K instead of 1M) compared with the state of the art methods. We experimentally demonstrated the ad hoc agent's ability to collaborate with other agents in complex environments, adapting to previously unknown changes (e.g., in agent types or team composition) to provide performance comparable with or better than the existing methods (Dodampegama and Sridharan 2023a,b).

The second example is changing-contact robot manipulation, which involves a robot making and breaking con-

tacts with different objects and surfaces; many robot and human manipulation tasks are such changing-contact tasks. The dynamics of these tasks are piecewise continuous, with abrupt transitions (i.e., sudden changes in force and acceleration) that can damage the robot or the domain objects. Unlike existing data-driven methods that attempt to explore different possible transitions in advance, and pose the problem of smooth motion as an offline optimization problem or learning problem (Khader et al. 2020), we drew inspiration from insights into human motor control (Flanagan et al. 2003; Kawato 1999). Specifically, we enabled the robot to rapidly learn and revise simple models that predict the end-effector sensor observations in the next step based on a single initial demonstration of the movement and run-time observations. During run-time, any mismatch between predicted and actual sensor measurements revises the predictive *forward model* and the gain parameters of a simple force-motion control law. Using experiments conducted in different simulation domains and on a physical robot manipulator, we demonstrated the ability to ensure smooth motion while performing changing-contact manipulation tasks with changes in surfaces and contacts that the robot was not aware of before (Sidhik, Sridharan, and Ruiken 2024).

Interactive learning for visual scene understanding and planning. In addition to the AHT example above, I present two examples to illustrate the benefits of leveraging the interplay between reasoning and learning in architectures that embed the outlined principles. I intentionally pick examples that illustrate the use of these principles in conjunction with modern data-driven AI systems.

The first example focuses on vision-based scene understanding, planning, and question answering, which are fundamental problems in computer vision and robotics. Methods considered to be state of the art for these problems are based on deep networks that are trained, for example, with a large dataset of images, potential questions, and answers to these questions. We, on the other hand, designed architectures based on the principles outlined above. Specifically, we adapted our refinement-based architecture to determine the occlusion of objects and the stability of object structures in images, arrange objects in desired configurations, and to answer questions about the decisions made. This architecture performed non-monotonic logical reasoning with generic domain knowledge available a priori to make the desired decisions (e.g., about stability and occlusion) if possible. Examples that could not be handled through reasoning triggered learning, with the robot automatically identifying such examples and the relevant regions in the corresponding images to be used for learning. Although this learning can be accomplished using one of many different methods, we intentionally used deep networks so that we could also use them as baselines for comparison. In addition, the corresponding training examples were also processed by an approach based on decision heuristics to induce new domain knowledge (e.g., actions and axioms) to be used for subsequent reasoning. We experimentally demonstrated: (a) performance comparable with or better than systems based just on deep networks, while using orders of magnitude fewer training examples; (b) faster training and better accuracy with deep networks when used as needed and only with relevant examples; and (c) performance improvement over time when reasoning and learning bootstrap off of each other (Riley and Sridharan 2019; Sridharan and Mota 2023). We also demonstrated that the architecture enables a robot to provide relational descriptions at different abstractions as explanations in response to different types of questions (causal, contrastive, counterfactual) (Mota, Sridharan, and Leonardis 2021; Sridharan 2024).

The second example illustrates the inclusion of an LLM in an architecture based on the outlined principles. Specifically, we developed an architecture that enabled an *embodied (AI) agent*[2] to collaborate with a human in completing assigned tasks in a home environment. Many papers have incorrectly claimed that LLMs can "plan" although their operation does not match the original interpretation of *planning*, as highlighted in a recent tutorial and position paper (Kambhampati et al. 2024). Instead, our architecture drew inspiration from work on *LLM-Modulo frameworks* (Guan et al. 2023), and used an LLM to provide a generic prediction about the sequence of tasks likely to be assigned in the near future. The AI agent considered these anticipated tasks as joint goals in conjunction with the current task. It then incorporated decision heuristics with classical planning methods to compute action sequences that would enable it to achieve these goals in collaboration with the human. We demonstrated substantial improvement in accuracy and computational efficiency of task completion compared with a system that only used an LLM or did not anticipate future tasks (Arora et al. 2024).

In summary, this paper is an attempt to promote appreciation for some fundamental principles that can be traced back to the early pioneers of AI but are not being fully leveraged in the design of modern AI systems. I hope that the examples provided in this paper will encourage researchers to incorporate these principles in the architectures they develop to address various open problems in robotics. More broadly, with the increasing development and use of AI methods in different disciplines such as Astrophysics, Chemistry, and Climate science, these principles and related observations are also relevant to research in AI and these other disciplines.

Acknowledgements

The ideas described in this paper were informed by research threads pursued in collaboration with Hasra Dodampegama, Michael Gelfond, Gerd Gigerenzer, Rocio Gomez, Konstantinos Katsikopoulos, Pat Langley, Ales Leonardis, Ben Meadows, Tiago Mota, Heather Riley, Saif Sidhik, Özgür Şimşek, Jeremy Wyatt, and Shiqi Zhang. This work was supported in part by the U.S. ONR Awards N00014-13-1-0766, N00014-17-1-2434, N00014-20-1-2390, and N00014-24-1-2737, AOARD award FA2386-16-1-4071, and U.K. EPSRC award EP/S032487/1. All conclusions described in this paper are those of the author alone.

[2] Although the phrase "embodied AI agent" is a reference to the fact that the agent senses and interacts with the domain, it is ironically often used to refer to such interaction in simulation domains instead of the physical world.

References

Arora, R.; Singh, S.; Swaminathan, K.; Banerjee, S.; Bhowmick, B.; Jatavallabhula, K. M.; Sridharan, M.; and Krishna, M. 2024. Anticipate & Act: Integrating LLMs and Classical Planning for Efficient Task Execution in Household Environments. In *International Conference on Robotics and Automation (ICRA)*. Yokohoma, Japan.

Banihashemi, B.; Giacomo, G. D.; and Lesperance, Y. 2017. Abstractions in Situation Calculus Action Theories. In *AAAI Conference on Artificial Intelligence*, 1048–1055. San Francisco, USA.

Banihashemi, B.; Giacomo, G. D.; and Lesperance, Y. 2018. Abstraction of Agents Executing Online and their Abilities in Situation Calculus. In *International Joint Conference on Artificial Intelligence*. Stockholm, Sweden.

Barrett, S.; Rosenfeld, A.; Kraus, S.; and Stone, P. 2017. Making friends on the fly: Cooperating with new teammates. *Artificial Intelligence*, 242: 132–171.

Besold, T. R.; d'Avila Garcez, A. S.; Bader, S.; Bowman, H.; Domingos, P. M.; Hitzler, P.; Kühnberger, K.-U.; Lamb, L. C.; Lowd, D.; Lima, P. M. V.; de Penning, L.; Pinkas, G.; Poon, H.; and Zaverucha, G. 2022. Neural-Symbolic Learning and Reasoning: A Survey and Interpretation. In Hitzler, P.; and Sarker, M. K., eds., *Neuro-Symbolic Artificial Intelligence: The State of the Art*. IOS Press, Amsterdam.

Brighton, H.; and Gigerenzer, G. 2012. How Heuristics Handle Uncertainty? In *Ecological Rationality: Intelligence in the World*. New York: Oxford University Press.

Broadbent, D. E. 1957. A Mechanical Model for Human Attention and Immediate Memory. *Psychology Review*, 64: 205–215.

Curtis, A.; Silver, T.; Tenenbaum, J. B.; Lozano-Perez, T.; and Kaelbling, L. P. 2022. Discovering State and Action Abstractions for Generalized Task and Motion Planning. In *AAAI Conference on Artificial Intelligence*, 5377–5384.

Dodampegama, H.; and Sridharan, M. 2023a. Back to the Future: Toward a Hybrid Architecture for Ad Hoc Teamwork. In *AAAI Conference on Artificial Intelligence*. Washington DC, USA.

Dodampegama, H.; and Sridharan, M. 2023b. Knowledge-based Reasoning and Learning under Partial Observability in Ad Hoc Teamwork. *Theory and Practice of Logic Programming*, 23(4): 696–714.

Doshi, R.; Walke, H.; Mees, O.; Dasai, S.; and Levine, S. 2024. Scaling Cross-Embodied Learning: One Policy for Manipulation, Navigation, Locomotion, and Aviation. In *International Conference on Robot Learning*. Munich, Germany.

Durbach, I. N.; Algorta, S.; Kantu, D. K.; Katsikopoulos, K. V.; and Simsek, O. 2020. Fast and Frugal Heuristics for Portfolio Decisions with Positive Project Interactions. *Decision Support Systems*, 138.

Flanagan, J. R.; Vetter, P.; Johansson, R. S.; and Wolpert, D. M. 2003. Prediction Precedes Control in Motor Learning. *Current Biology*, 13: 146–150.

Freeman, T.; and Pfenning, F. 1991. Refinement Types for ML. In *ACM SIGPLAN Conference on Programming Language Design and Implementation*, 268–277. Toronto, Canada.

Garrett, C. R.; Chitnis, R.; Holladay, R.; Kim, B.; Silver, T.; Kaelbling, L. P.; and Lozano-Perez, T. 2021. Integrated Task and Motion Planning. *Annual Review of Control, Robotics, and Autonomous Systems*, 4: 265–293.

Gigerenzer, G. 2016. *Towards a Rational Theory of Heuristics*, 34–59. London: Palgrave Macmillan UK.

Gigerenzer, G. 2020. What is Bounded Rationality? In *Routledge Handbook of Bounded Rationality*. Routledge.

Gigerenzer, G. 2021. Embodied Heuristics. *Frontiers in Psychology*, 12: 1–12.

Gigerenzer, G.; and Gaissmaier, W. 2011. Heuristic Decision Making. *Annual Review of Psychology*, 62: 451–482.

Gomez, R.; Sridharan, M.; and Riley, H. 2021. What do you really want to do? Towards a Theory of Intentions for Human-Robot Collaboration. *Annals of Mathematics and Artificial Intelligence, special issue on commonsense reasoning*, 89: 179–208.

Guan, L.; Valmeekam, K.; Sreedharan, S.; and Kambhampati, S. 2023. Leveraging Pre-trained Large Language Models to Construct and Utilize World Models for Model-based Task Planning. In *International Conference on Neural Information Processing Systems*. New Orleans, USA.

Huang, W.; Xia, F.; Shah, D.; Driess, D.; Zeng, A.; Lu, Y.; Florence, P. R.; Mordatch, I.; Levine, S.; Hausman, K.; and Ichter, B. 2023. Grounded Decoding: Guiding Text Generation with Grounded Models for Robot Control. In *International Conference on Neural Information Processing Systems*. New Orleans, USA.

Kambhampati, S.; Valmeekam, K.; Guan, L.; Verma, M.; Stechly, K.; Bhambri, S.; Saldyt, L. P.; and Murthy, A. 2024. Position: LLMs Can't Plan, But Can Help Planning in LLM-Modulo Frameworks. In *International Conference on Machine Learning*. Vienna, Austria.

Katsikopoulos, K.; Simsek, O.; Buckmann, M.; and Gigerenzer, G. 2021a. *Classification in the Wild: The Science and Art of Transparent Decision Making*. MIT Press.

Katsikopoulos, K.; Simsek, O.; Buckmann, M.; and Gigerenzer, G. 2021b. Transparent modeling of influenza incidence: Big data or a single data point from psychological theory? *International Journal of Forecasting*.

Kawato, M. 1999. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, (6): 718–727.

Khader, S. A.; Yin, H.; Falco, P.; and Kragic, D. 2020. Data-efficient model learning and prediction for contact-rich manipulation tasks. *IEEE Robotics and Automation Letters*, 5(3): 4321–4328.

Laird, J. E.; Gluck, K.; Anderson, J.; Forbus, K. D.; Jenkins, O. C.; Lebiere, C.; Salvucci, D.; Scheutz, M.; Thomaz, A.; Trafton, G.; Wray, R. E.; Mohan, S.; and Kirk, J. R. 2017. Interactive Task Learning. *IEEE Intelligent Systems*, 32(4): 6–21.

Langley, P.; and Katz, E. P. 2022. Motion Planning and Continuous Control in a Unified Cognitive Architecture. In *Annual Conference on Advances in Cognitive Systems*. Arlington, VA.

Langley, P.; Meadows, B.; Sridharan, M.; and Choi, D. 2017. Explainable Agency for Intelligent Autonomous Systems. In *Innovative Applications of Artificial Intelligence*. San Francisco, USA.

Langley, P.; Sridharan, M.; and Meadows, B. 2018. Representation, Use, and Acquisition of Affordances in Cognitive Systems. In *AAAI Spring Symposium on Integrating Representation, Reasoning, Learning and Execution for Goal Directed Autonomy*. Stanford, USA.

Lovas, W.; and Pfenning, F. 2010. Refinement Types for Logical Frameworks and their Interpretation as Proof Irrelevance. *Logical Methods in Computer Science*, 6(4).

Mellies, P.-A.; and Zeilberger, N. 2015. Functors are Type Refinement Systems. In *ACM SIGPLAN-SIGACT Symsposium on Principles of Programming*, 3–16. Mumbai, India.

Minsky, M. L. 1986. *The Society of Mind*. Simon and Schuster.

Mirsky, R.; Carlucho, I.; Rahman, A.; Fosong, E.; Macke, W.; Sridharan, M.; Stone, P.; and Albrecht, S. 2022. A Survey of Ad Hoc Teamwork: Definitions, Methods, and Open Problems. In *European Conference on Multiagent Systems*.

Moravec, H. P. 1990. *Mind Children: The Future of Robot and Human Intelligence*. Harvard University Press.

Mota, T.; Sridharan, M.; and Leonardis, A. 2021. Integrated Commonsense Reasoning and Deep Learning for Transparent Decision Making in Robotics. *Springer Nature CS*, 2(242): 1–18.

Rahman, M. A.; Hopner, N.; Christianos, F.; and Albrecht, S. V. 2021. Towards Open Ad Hoc Teamwork Using Graph-based Policy Learning. In *International Conference on Machine Learning*, 8776–8786.

Riley, H.; and Sridharan, M. 2019. Integrating Non-monotonic Logical Reasoning and Inductive Learning With Deep Learning for Explainable Visual Question Answering. *Frontiers in Robotics and AI, special issue on Combining Symbolic Reasoning and Data-Driven Learning for Decision-Making*, 6: 20.

Santos, P. M.; Ribeiro, J. G.; Sardinha, A.; and Melo, F. S. 2021. Ad Hoc Teamwork in the Presence of Non-stationary Teammates. In Marreiros, G.; Melo, F. S.; Lau, N.; Lopes Cardoso, H.; and Reis, L. P., eds., *Progress in Artificial Intelligence*, 648–660. Springer International.

Schick, T.; Dwivedi-Yu, J.; Dessi, R.; Raileanu, R.; Lomeli, M.; Zettlemoyer, L.; Cancedda, N.; and Scialom, T. 2023. Toolformer: Language Models Can Teach Themselves to Use Tools. In *Advances in Neural Information Processing Systems*. New Orleans, USA.

Sidhik, S.; Sridharan, M.; and Ruiken, D. 2024. An Adaptive Framework for Trajectory Following in Changing-Contact Robot Manipulation Tasks. *Robotics and Autonomous Systems*, 181: 1–21.

Simon, H. A. 1955. A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 69: 99–118.

Simon, H. A. 1956. Rational Choice and the Structure of the Environment. *Psychological Review*, 63: 129–138.

Sloman, A. 2012. Meta-morphogenesis and the Creativity of Evolution. In *Workshop on Computational Creativity, Concept Invention, and General Intelligence at ECAI*. Montpellier, France.

Smet, L. D.; Martires, P. Z. D.; Manhaeve, R.; Marra, G.; Kimmig, A.; and Readt, L. D. 2023. Neural Probabilistic Logic Programming in Discrete-Continuous Domains. In *International Conference on Uncertainty in Artificial Intelligence*, 529–538.

Sridharan, M. 2024. Integrated Knowledge-based Reasoning and Data-driven Learning for Explainable Agency in Robotics. In *Explainable Agency in Artificial Intelligence: Research and Practice*. CRC Press.

Sridharan, M.; Gelfond, M.; Zhang, S.; and Wyatt, J. 2019. REBA: A Refinement-Based Architecture for Knowledge Representation and Reasoning in Robotics. *Journal of Artificial Intelligence Research*, 65: 87–180.

Sridharan, M.; and Meadows, B. 2019. Towards a Theory of Explanations for Human-Robot Collaboration. *Kunstliche Intelligenz*, 33(4): 331–342.

Sridharan, M.; Meadows, B.; and Gomez, R. 2017. What can I not do? Towards an Architecture for Reasoning about and Learning Affordances. In *International Conference on Automated Planning and Scheduling*. Pittsburgh, USA.

Sridharan, M.; and Mota, T. 2023. Towards Combining Commonsense Reasoning and Knowledge Acquisition to Guide Deep Learning. *Autonomous Agents and Multi-Agent Systems*, 37(4).

Srivastava, S.; Riano, L.; Russell, S.; and Abbeel, P. 2013. Using Classical Planners for Tasks with Continuous Operators in Robotics. In *International Conference on Automated Planning and Scheduling (ICAPS)*. Rome, Italy.

Surís, D.; Menon, S.; and Vondrick, C. 2023. ViperGPT: Visual Inference via Python Execution for Reasoning. In *IEEE International Conference on Computer Vision (ICCV)*. Paris, France.

Triesman, A. M.; and Gelade, G. 1980. A Feature-Integration Theory of Attention. *Cognitive Psychology*, 12: 97–136.

Turing, A. 1952. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London B*, 237(641): 37–72.

Zhang, H.; Du, W.; Shan, J.; Zhou, Q.; Du, Y.; Tenenbaum, J. B.; Shu, T.; and Gan, C. 2024. Building Cooperative Embodied Agents Modularly with Large Language Models. In *International Conference on Learning Representations*.

Zhao, Z.; Lee, W. S.; and Hsu, D. 2023. Large Language Models as Commonsense Knowledge for Large-Scale Task Planning. In *International Conference on Neural Information Processing Systems*.