The Problem with Problems

Michael T. Cox MICHAEL.COX@WRIGHT.EDU
Wright State Research Institute, Wright State University, Dayton, OH 45435 USA

Abstract

For over 50 years the artificial intelligence community has represented problems to be solved as a combination of an initial state and goal state along with some background knowledge. In this paper, I challenge this representation because it does not adequately capture the nature of a problem. Instead, a problem is a state of the world that limits choice in terms of potential goals or available actions. To begin to capture this view of problems, a representation must include a characterization of the context that exists when a problem arises and an explanation that causally links the part of the context that contributes to the problem with the goal that constitutes a solution. The challenge to the AI community is to not only represent such features but to enable agents that can infer them autonomously.

1. Introduction

The task of problem-solving was a central process examined during the genesis of the field of artificial intelligence (AI). Like humans, a machine should be capable of solving difficult problems if it were to be considered intelligent. To illustrate such behavior, programs such as the *General Problem Solver (GPS)* were given some initial starting state and a goal state and then they would output a sequence of actions that would achieve the goal (Newell & Simon, 1963). This sequence of actions (i.e., plan) was considered a solution to the problem. Problem-solving itself was cast as a search through the state-space implicit in its knowledge (in the case of GPS, inherent in its difference table) to find a combination of steps that meet the goal criteria (Amarel, 1968; McCarthy & Hayes, 1969).

Over the years, many types of problems have been studied. Initially, researchers developed algorithms for various puzzles and games such as the Towers of Hanoi[1] (e.g., Ernst, 1969; Knoblock, 1990), chess (e.g., Bilalić, McLeod, & Gobet, 2008; Chase & Simon, 1973; Hsu, 2002), and the 8-puzzle and its derivations (e.g., Ratner, & Warmuth, 1986; Russell, & Norvig, 2003). As research matured, research turned to design and planning tasks. For design problems, solutions are artifact design configurations that meet specific requirements (Chandrasekaran, 1990; Maher, Balachandran, & Zhang, 1995); for planning, solutions are sequences of steps that achieve a goal (Ghallab, Nau, & Traverso, 2004). This paper will focus on planning problems to illustrate the argument in some depth. Further, we will attempt to dispel the notion that puzzles

[1] At least 340 articles were published on the game in the 100 years from its invention in 1883-1983 (Stockmeyer 2013), and apparently even ants can solve a isomorphic version of the problem (Reid, Sumpter, & Beekman, 2010).

are problems in the sense of the word whereby problems imply risk for an agent. Puzzles in isolation do not contain threat or risk or in any way limit the choices available to an agent as do problems. The defining attribute of a problem is the restriction of choice.

This paper has two major sections. The first section outlines the classical representation of a problem and then enumerates some problems with this construction. The second section proposes an alternative problem representation and then challenges the AI community to take serious the computational tasks of recognizing a problem in any state, explaining what causes it, and autonomously generating goals to remove these causes.

2. The Classical Problem Definition

What is a problem? Answer: An initial state, the goal state, and the means to get from one to the other.

2.1 Classical Problem Representation

Over time, the representation of a problem has been formalized with a standard notation. Here we use the formalism from the automated planning community (e.g., Bonet & Geffner, 2001; and especially Ghallab, Nau, & Traverso, 2004), but variations across AI tend to be similar to the following representations.

Formal Problem Definition: A problem consists of an initial state, goal expression, and a domain transition model.

$$\mathcal{P} = (s_0, g, \Sigma) \text{ where } s_0 \in S, g \in G \subset S$$

State Transition System: a set of possible states, a set of actions, and a successor function.

$$\Sigma = (S, A, \gamma)$$

Successor Function: returns the next state, given a current state and action.

$$\gamma : S \times A \rightarrow S$$

Problem Solution: a sequence of actions, i.e., a plan.

$$\pi : 2^A = \alpha_1 \mid \pi[2 \ldots n] = \langle \alpha_1, \alpha_2 \ldots \alpha_n \rangle$$

Plan Execution: starting from the initial state (s_0), recursive action execution results in the goal state (s_g) that entails the goal expression.

$$\gamma(s_0, \pi) = \gamma(\gamma(s_0, \alpha_1), \pi[2 \ldots n]) \rightarrow s_g \models g$$

2.2 Problems with the Classical Representation

The biggest problem with the classical representation is that problems in this form amount to arbitrary states to achieve, i.e., puzzles, not problems. The reason that the initial state is problematic and the goal state not is lacking in the representation. Instead the causal justification remains in the head of the researcher; the machine has no access to it and thus must simply do what it is told.

Classical problems therefore tend to possess three negative attributes as follows.

- The need for the goal or why it is a solution is opaque and cannot be explained;
- The central issue with the initial state is implicit;
- Problems must be provided by humans rather than inferred by an intelligent system or agent.

The representation for problems is often overly simplified in the literature. Consider the blocksworld planning domain (Gupta & Nau, 1992; Winograd, 1972). Initial states in this domain are random configurations of blocks, and so too are the goals. For example in the first panel of Figure 1, the initial state is the arrangement of three blocks on the table and the goal state is to have block A on top of block B. The planner blindly executes a plan to pick up A and stack it on B, but the planner has no reason why this goal state is valued, and if the world changes dramatically the agent can only adapt the plan to maintain the intent without cause.

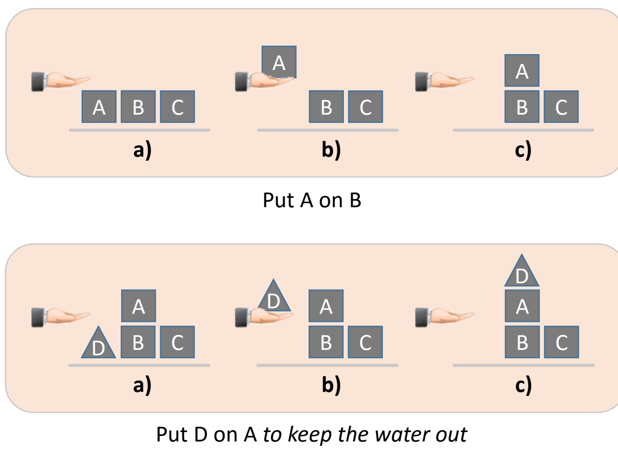

Figure 1. Blocksworld state sequences that distinguish justified problems from arbitrary problems (adapted from Cox, 2013).

In the second panel, we assume a larger context such as the construction of buildings. In this context, the planner wishes to have the triangle D on the block A to keep water out when it rains. Here D represents the roof of the house composed of A, B, and C. Water getting into a person's living space is a problem; stacking random blocks is not.

In both cases, papers that have been conditionally accepted will be examined for compliance with conditions placed on their acceptance. Final papers must follow the content, format, and length restrictions that are specified in these instructions.

3. An Alternative Problem Definition

What is a problem? Answer: A situation that limits choice in terms of potential goals or available actions.

Problems are not simply puzzles or arbitrary states to achieve. A problem is a situation relative to an agent (or agents) with some existing history of actions and decisions.2 We claim that a situation is a problem for an agent whenever a significant risk exists (either immediately or eventually) of a loss in ability to achieve its current or future goals or to select and execute particular actions.

Potential goals are those that might be possible to formulate in the future; kinetic goals are those currently in an agent's agenda. Risks to either can pose problems. For example, the loss of home value due to neighborhood trends is a problem for a house's owner. It limits the potential goal of having the house sold, even if the owner does not currently have the desire to do so.

Formal Problem Definition: A problem consists of the currently observed and expected states, the background knowledge, an episodic history, a causal explanation of the problem and a goal that solves it. We examine these in turn.

$$\mathcal{P} = (s_c,\ s_e,\ Bk, H_c,\ \chi, g')\ where\, s_c, s_e \in S$$

3.1 Representing the Intent Context

The key to understanding problems is to recognize the importance of goals or the intended future directions of an agent. A new area of research called *goal reasoning* (or sometimes referred to as *goal-driven autonomy*) has attempted to develop agents with a capability to reason about their own goals, to change them when warranted, and to formulate new goals when confronted with new problems (Aha 2018;[3] Cox 2007; 2013; Klenk, Molineaux & Aha, 2013; Munoz-Avila 2018). To do so, problems must include a representation of the dynamic context of the agent with respect to its intent.

Background Knowledge: the state transition system (see Classical Problem Representation) along with a set of goal operations.

$$Bk = (\Sigma,\ \Delta)$$

Here the action models within Σ enable an agent to predict subsequent states (s_e) and to use these expectations in comparison with observed states (s_c) to suspect the presence of problems. See Dannenhauer & Munoz-Avila (2015; Dannenhauer, Munoz-Avila & Cox, 2016) for detail.

Interpretation Function: given a state and a (possibly empty) goal, the interpretation function performs goal operations from $\Delta = \{\delta \mid \delta : G \rightarrow G\}$ resulting in a desired goal (Cox, 2017; Cox, Dannenhauer, & Kondrakunta, 2017).

$$\beta : S \times G \rightarrow G$$

A specific operation from Δ is represented as the 4-tuple $\delta = (head(\delta),\ parameter(\delta),$ $pre(\delta),\ res(\delta)$, where $pre(\delta)$ and $res(\delta)$ are its preconditions and result. The transformation's identifier is $head(\delta)$, and its input goal argument is $parameter(\delta)$. There are two essential goal operations. *Goal formulation* ($\beta(s, \emptyset) \rightarrow g$) infers a new goal given some state (Cox, 2007; 2013;

[2] The idea of an initial state is itself arbitrary, if one allows for agents that persist over extended time periods.

[3] Based on the Robert S. Engelmore Memorial Lecture given at IAAI-17.

Paisner, Maynord, Cox & Perlis, 2013); whereas, *goal change* ($\beta(s, g) \to g'$) transforms an existing goal into another (Choi, 2011; Cox & Veloso, 1998; Cox & Dannenhauer, 2016).[4]

Goal Trajectory: the original goal (g_1) and its evolution into current goal (g_c).

$$\vec{g} = \langle\!\langle (s_0,\ g_1), (s_i, \beta(s_i, g_1)), \dots (s_j, g_c) \rangle\!\rangle$$

Goals do not always remain as given or first formulated. They are malleable objects that change over time as agents change their intent. Goals go through arcs or trajectories in a goal hyperspace over time (see Bengfort & Cox, 2015; Eyorokon, Panjala, & Cox, 2017; Eyorokon, Yalamanchili, & Cox, 2018).

Current Goal Agenda: the current goal being solved along with any pending goals remaining to be solved.

$$\hat{G}_c = \{g_x, g_y, \dots g_c\}$$

Agenda History: the evolution of the agenda up to and including its current instance.

$$\hat{G}_h = \langle \hat{G}_1, \hat{G}_2, \dots \hat{G}_c \rangle$$

3.2 Representing the Problem-Solving Episode

Finally, the problem representation requires a formalism for the problem-solving process and its unfolding solutions to a goal trajectory. The reason for this requirement is that new problems can arise during the act of solving a previous problem or during plan execution in the world.

Plan: previously executed steps (including current step $\alpha_{\boldsymbol{c}}$) composed with all remaining steps ($\pi_{\boldsymbol{r}}$).

$$\pi\colon 2^A = \langle \alpha_{\mathbf{1}}, \alpha_{\mathbf{2}}, \dots \alpha_{\boldsymbol{c}} \rangle \circ \pi_{\boldsymbol{r}} = \pi_{\boldsymbol{c}} \circ \pi_{\boldsymbol{r}}$$

(Re)Planning Function: given a state, a goal, and a (possibly empty) plan, the planning function performs a (re)planning operation using Σ (Cox 2017).

$$\varphi\colon S \times G \times 2^A \to 2^A$$

Traditional plan generation is of the following form.

$$\varphi(s_0, g_1, \emptyset) \to \pi_1$$

Planning Trajectory: the sequence over time of changing plans paired with the goals they purport to solve from the first goal and plan (g_1, π_1) until and including the current goal (g_c) where the remainder of the plan (π_r) awaits execution.

$$\vec{\pi} = \langle (g_1, \pi_1), \Big(g_i, \varphi\big(s_j, g_i, \pi_1[k \dots n]\big)\Big), \dots (g_c, \pi_r) \rangle$$

Current Execution Episode: the sequence of states and executed actions that occurred before and not including the current state.

---

[4] Goal formulation is implemented as the *insertion transformation* $\delta^*(\emptyset) = g$; whereas, a goal change would be the *identity transformation* $\delta^I(g_i) = g_i$ for all $g_i \in G$ i.e., the tuple $(identity, g, \{true\}, g)$. See Cox (2017) for further detail and Cox & Dannenhauer (2017) for a more expressive goal representation.

$$\varepsilon_c = \langle s_0, \alpha_1, \gamma(s_0, \alpha_1), \alpha_2, \ldots s_{c-1}, \alpha_c \rangle$$

Episodic History: the history includes the goal, agenda, plan, and execution trajectories.

$$H_c = (\vec{g}, \hat{G}_h, \vec{\pi}, \varepsilon_c)$$

3.3 The AI Challenge: Inferring the Problem

Finally, we have the constituents to fully specify a problem and the restriction of choice it represents for an agent. If an agent is building a physical structure to contain its possessions and to safely house itself, it will have a typical set of goals to complete and reasons for each. The goal to add the roof is causally connected to the need for guarding one's possessions and for personal safety and comfort. However, these ancillary goals are not currently threatened at the current time given the possessions are safe elsewhere, it is not raining, and the agent does not currently live in the house. But if possessions are moved into the house and a proper roof is not in place, the possessions will lose value when it rains. Lost value signifies reduced benefit and therefore less choice. This explanation (or others like it relating the current state to what can occur in the future) supports the goal of having a roof placed on the structure. Such relationships become institutionalized in best practices and behavior (e.g., building codes or procedures), but they are crucial in relatively novel situations that represent new problems to an agent.

Problem Explanation: an explanatory graph of vertices and edges causally linking the current state (s_c) to the limitation of choice.

$$\chi = (V, E)$$

New research in goal reasoning has argued that an intelligent agent should not simply solve problems and achieve goals presented to it, rather it should be capable of (1) recognizing problems on their own; (2) explaining what caused them; and (3) generating an independent goal to solve the problem or remove the cause (Cox, 2013). Preliminary findings show some benefit to this approach, although it is quite difficult to automatically separate out "true" problems from minor anomalies or discrepancies encountered by an agent in a changing, complex environment (Kondrakunta et al., 2018; Gogineni, Kondrakunta, Molineaux, & Cox, 2018).[5]

This is the next big challenge for the AI community in my opinion. AI systems or intelligent agents, if they are to be genuinely autonomous with a significant measure of independence, should themselves infer χ and g' (placing the latter in $\hat{G}_c$). They should not simply generate π and then wait for a human to given them further direction (or worse, halt computation). By this challenge, a problem would be represented as the following reduced 4-tuple. Neither the goal nor the explanation would be given.

$$\mathcal{P} = (s_c, s_e, Bk, H_c)$$

A solution to $\mathcal{P}$ would thus be of the form (χ, g', π). If we prevail over time in this task, the effort will result in agents that have a flexible capacity for dealing with problems on their own

[5] These publications also illustrate the approach with a much more complex explanation and further distinctions that size restrictions in this paper do not allow. The formal representation developed here is not as well described in the papers however.

and if necessary explaining to others the reasons for their choices using the kind of intelligence reserved until now by only humans.

4. Conclusion

The planning community is beginning to open up to the view that planning agents are more than generators of sequences of action steps; they must consider online, dynamic, and uncertain environments where decisions, action execution, interaction with other agents, and replanning all combine (Ghallab, Nau, & Traverso, 2014; 2016).However, the representation of an actual problem remains much the same as it has for fifty years (e.g., Patra, Traverso, Ghallab, & Nau, 2018). Goal reasoning and *explainable AI* (Aha, et al., 2017; Cox, 1994; 2011; Gunning, 2016; Lane, Core, van Lent, Solomon, & Gomboc, 2005) are two important new areas of research with growing interest that question the status quo and push the frontiers of what we think machines should be able to accomplish on their own.

Acknowledgements

I thank Cindy Marling for her comments and insights with regard to an earlier draft of this paper. The material is based upon work supported by AFOSR grant #FA2386-17-1-4063, by ONR grant #N00014-18-1-2009, and by DARPA contract number N6600118C4039.

References

Aha, D. W. (2018). Goal reasoning: Foundations, emerging applications, and prospects. *AI Magazine 39*(2), 3-24.

Aha, D. W., Darrell, T., Pazzani, M., Reid, D., Sammut, C., & Stone, P. (Eds.) (2017). *Proceedings of the IJCAI-17 Workshop on Explainable AI (XAI)*. IJCAI.

Amarel, S. (1968). On representations of problems of reasoning about actions. In D. Michie (Ed.), *Machine Intelligence 3* Edinburgh University Press.

Bengfort, B., & Cox, M. T. (2015). Interactive reasoning to solve knowledge goals. D. W. Aha (Ed.), *Goal reasoning: Papers from the ACS workshop* (pp. 10-25). Tech. Rep. No. GT-IRIM-CR-2015-001. Atlanta: Georgia Institute of Technology.

Bilalić, M., McLeod, P., & Gobet, F. (2008) Expert and "novice" problem solving strategies in chess: Sixty years of citing de Groot (1946), *Thinking & Reasoning, 14*:4, 395-408.

Bonet, B., & Geffner, H. (2001). Planning as heuristic search. *Artificial Intelligence*, *129*, 5–33.

Chandrasekaran, B. (1990). Design problem solving: A task analysis *AI Magazine 11*(4). 59-71.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.

Choi, D. (2011). Reactive goal management in a cognitive architecture. *Cognitive Systems Research*, *12*, 293-308.

Cox, M. T. (1994). Machines that forget: Learning from retrieval failure of mis-indexed explanations. In A. Ram & K. Eiselt (Eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society* (pp. 225-230). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cox, M. T. (2007). Perpetual self-aware cognitive agents. *AI Magazine*, *28*(1), 32-45.

Cox, M. T. (2011). Metareasoning, monitoring, and self-explanation. In M. T. Cox & A. Raja (Eds.) *Metareasoning: Thinking about thinking* (pp. 131-149). Cambridge, MA: MIT Press.

Cox, M. T. (2013). Question-based problem recognition and goal-driven autonomy. In D. W. Aha, M. T. Cox, & H. Munoz-Avila (Eds.), *Goal Reasoning: Papers from the Advances in Cognitive Systems Workshop* (pp. 10-25). Tech. Rep. No. CS-TR-5029. College Park, MD: University of Maryland, Department of Computer Science.

Cox, M. T. (2017). A model of planning, action, and interpretation with goal reasoning. *Advances in Cognitive Systems, 5*, 57-76.

Cox, M. T., & Dannenhauer, D. (2016). Goal transformation and goal reasoning. In *Proceedings of the 4th Workshop on Goal Reasoning*. New York, IJCAI-16.

Cox, M. T., Dannenhauer, D., & Kondrakunta, S. (2017). Goal operations for cognitive systems. In *Proceedings of the Thirty-first AAAI Conference on Artificial Intelligence* (pp. 4385-4391). Palo Alto, CA: AAAI Press.

Cox, M. T., & Dannenhauer, Z. A. (2017). Perceptual goal monitors for cognitive agents in changing environments. In *Proceedings of the Fifth Annual Conference on Advances in Cognitive Systems, Poster Collection* (pp. 1-16). Palo Alto, CA: Cognitive Systems Foundation.

Cox, M. T., & Veloso, M. M. (1998). Goal transformations in continuous planning. In M. desJardins (Ed.), *Proceedings of the 1998 AAAI Fall Symposium on Distributed Continual Planning* (pp. 23-30). Menlo Park, CA: AAAI Press / MIT Press.

Dannenhauer, D., & Munoz-Avila, H. (2015). Raising expecta-tions in GDA agents acting in dynamic environments. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (pp. 2241–2247). Buenos Aires, Argentina.

Dannenhauer, D., Munoz-Avila, H., & Cox, M. T. (2016). Informed expectations to guide GDA agents in partially observable environments. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 2493–2499). New York.

Ernst, G. W. (1969). Sufficient conditions for the success of GPS. *Journal of the Association for Computing Machinery, 16*(4):517-533.

Eyorokon, V., Panjala, U., & Cox, M. T. (2017). Case-based goal trajectories for knowledge investigations. V. Rus & Z. Markov (Eds.), *Proceedings of the Thirtieth International FLAIRS Conference* (pp. 477-482). Palo Alto, CA: AAAI Press.

Eyorokon, V., Yalamanchili, P., & Cox, M. T. (2018). Tangent recognition and anomaly pruning to trap off-topic questions in conversational case-based dialogues. In M. Cox, P. Funk, & S. Begum (Eds.), *Case-Based Reasoning Research and Development: Proceedings of the 26th International Conference* (pp.213-228). Berlin: Springer.

Ghallab, M., Nau, D., & Traverso, P. (2004). *Automated planning: Theory and practice*. San Francisco: Morgan Kaufmann.

Ghallab, M., Nau, D., & Traverso, P. (2014). The actor's view of automated planning and acting: A position paper. *Artificial Intelligence, 208*, 1-17.

Ghallab, M., Nau, D., & Traverso, P. (2016). *Automated Planning and Acting*. Cambridge: Cambridge University Press.

Gogineni, V., Kondrakunta, S., Molineaux, M., & Cox, M. T. (2018). Application of case-based explanations to formulate goals in an unpredictable mine clearance domain. In M. Minor (Ed.), *26th International Conference on Case-Based Reasoning: Workshop Proceedings - Case-based Reasoning for the Explanation of Intelligent Systems* (pp. 42-51). ICCBR-18.

Gunning, D. (2016). *Explainable Artificial Intelligence (XAI)*. DARPA-BAA-16-53. Washington, DC: US Department of Defense.

Gupta, N., & Nau, D. (1992). On the complexity of blocks-world planning. *Artificial Intelligence. 56*: 223–254.

Hsu, F.-H. (2002). *Behind Deep Blue: Building the computer that defeated the world chess champion*. Princeton: Princeton University Press.

Klenk, M., Molineaux, M., & Aha, D. W. (2013). Goal-driven autonomy for responding to unexpected events in strategy simulations. *Computational Intelligence, 29*.

Knoblock, C. A. (1990). Abstracting the Tower of Hanoi. In *Working Notes of AAAI-90 Workshop on Automatic Generation of Approximations and Abstractions* (pp. 13-23).

Kondrakunta, S., Gogineni, V. R., Molineaux, M., Munoz-Avila, H., Oxenham, M., & Cox, M. T. (2018). Toward problem recognition, explanation and goal formulation. In *Working Notes of the 2018 IJCAI Goal Reasoning Workshop*. IJCAI.

Maher, M. L., Balachandran, M. B., & Zhang, D. M. (1995). *Case-based reasoning in design*. Psychology Press.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence. 4,* 463-502.

Munoz-Avail, H. (2018). Adaptive goal-driven autonomy. In M. T. Cox, P. Funk, & S. Begum (Eds.), *Case-Based Reasoning Research and Development: Proceedings of the 26th International Conference* (pp. 1-10). Berlin: Springer.

Newell, A., & Simon, H. A. (1963). GPS, a program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 279-293). New York: McGraw Hill.

Paisner, M., Maynord, M., Cox, M. T., & Perlis, D. (2013). Goal-driven autonomy in dynamic environments. D. W. Aha, M. T. Cox, & H. Munoz-Avila (Eds.), *Goal Reasoning: Papers from the ACS workshop* (pp. 79-94). Tech. Rep. No. CS-TR-5029. College Park, MD: University of Maryland, Department of Computer Science.

Patra, S., Traverso, P., Ghallab, M., & Nau, D. (2018). Controller synthesis for hierarchical agent interactions. In *Proceedings of the Sixth Annual Conference on Advances in Cognitive Systems, Poster Collection*. Palo Alto, CA: Cognitive Systems Foundation.

Ratner, D., & Warmuth, M. K. (1986). Finding a shortest solution for the n × n extension of the 15-puzzle is intractable. *National Conference on Artificial Intelligence*, Menloa Park, CA: AAAI Press.

Reid, C.R., Sumpter, D.J.T., & Beekman, M. (2011). Optimization in a natural system: Argentine ants solve the Towers of Hanoi. *Journal of Experimental Biology 214*:50-58.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach*, 2nd ed. Prentice Hall.

Stockmeyer, P. (2013). *The Tower of Hanoi: A bibliography*. researchgate.net/publication/250056612_The_Tower_of_Hanoi_A_Bibliography

Lane, H. C., Core, M. G., van Lent, M., Solomon, S., & Gomboc, D. (2005). Explainable artificial intelligence for training and tutoring. In C.-K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.), *Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology* (pp. 762-764). Amsterdam: IOS Press.

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology, 3*(1), 1-19