Cognitive Design for Artificial Minds

Antonio Lieto

COGNITIVE DESIGN FOR ARTIFICIAL MINDS

by

Antonio Lieto

Available on the main online bookshops:

- Amazon: https://www.amazon.com/gp/product/1138207950/
- Walmart: https://www.walmart.com/ip/Cognitive-Design-for-Artificial-Minds-Paperback-9781138207950/571399366
- Routledge: https://www.routledge.com/Cognitive-Design-for-Artificial-Minds/Lieto/p/book/9781138207950
- Taylor & Francis (e-book): https://doi.org/10.4324/9781315460536

Contents

List of Figures and Tables

Introduction

This book is about (re)building a bridge between two different "sciences of the artificial": Artificial Intelligence and Cognitive Science that, nowadays, apart from some notable exceptions, do not talk much to each other as they should. Here, I review some of the main themes that have characterized the historical paths of these two disciplines and argue that the technological maturity reached in several domains now calls for a renewed joint enterprise finalised at addressing more substantial challenges that these two disciplines have to face from a scientific viewpoint.

The book explicitly targets a multidisciplinary audience. As such it is mainly an act of courage (or probably of irresponsibility) since experts in the specific subfields will have for sure much more things to say and would surely be able to communicate their own ideas in a better way than I can do. However, as mentioned, this book privileges the breath of the connections between the disciplines rather than the depths of the exploration within each single discipline. As such, it is not a manual or a handbook since it presupposes the knowledge of same basic elements of each of the disciplines that will be touched by our arguments. Of course, scholars and students of the diverse fields have knowledge of different pieces of the entire puzzle and need to be briefly introduced to the aspects that they do not know. This service is provided in the book that, however, reminds to the specialistic literature for the details.

One of the main goals of this manuscript is to show to the reader that the so called "cognitive design approach" has still an important role to play in the development of intelligent AI technologies as well as in the context of development of plausible computational models of cognition. In other words: the study of the "Cognitive Design" principles for building "Artificial Minds" will be hopefully a useful instrument for the current and future generation of AI and cognitive science scholars and students. In this respect, a first *caveat* is necessary: in the philosophical literature on the AI there are many different, and well known, positions about whether or not it it is justifiable to use the terms "mind", "intelligence" or "thinking", to describe the constitutive or the behavioural elements of a computational system. In this book we will not enter in the details of such a monumental and decades-long debate, that also involve the attribution of such faculties to other "species" (from non-human mammals to bacteria). Given the actual purpose of the book, we will also avoid to roughly summarize it because such an attempt would be necessarily incomplete. Sometimes, however, we will refer to some instances of such a debate. For the moment we just mention here, as a reference for the position about why the term "mind" can be justifiably

associated to the term "artificial", the book "Artificial Minds" by Stan Franklin (Franklin, 1995). The position defended by Franklin, that sees the possession of a "mind" as a matter of degrees and not as a mere boolean notion and that - as such - foresee the possibility of implementing (to some degree) a "mind" in an artificial system, can be considered our starting working hypothesis.

A Paola e Francesca

Chapter 1 - **Cognitive Science and Artificial Intelligence: Death and Rebirth of a collaboration**

The first chapter proposes a brief historical overview of some of the main insights developed in 65 years of research in Artificial Intelligence (AI), by introducing the early vision of the discipline (based on a mutual collaboration with the Cognitive Psychology) and its "paradigm shift" started from the mid'80's of the last century. Starting from that period, Artificial Intelligence and the interdisciplinary enterprise known as Cognitive Science started to produce several sub-fields, each with its own goals, methods and criteria for evaluation. The reasons for the current renewed interest of a cognitively inspired approach in the AI research are discussed.

Chapter 2 - **Cognitive and Machine oriented Approaches to Intelligence in Artificial Systems**

This chapter presents the different possible routes to build an Artificial Intelligence system. On one hand it presents the design assumptions underlying the cognitive approaches to AI and, on the other hand, it presents the tenets of machine oriented approaches aimed at obtaining AI systems able to exhibit intelligent behaviour without making any assumption about the biological or cognitive plausibility of the implemented mechanisms. It additionally introduces the reader to the main instances about the debate on the levels of analysis of computational systems (being cognitively inspired or not).

Chapter 3 – **Principles of the Cognitive Design Approach**

This chapter introduces the classical notions of rationality developed in the field of cognitive modelling and presents different types of explanatory accounts available in the literature. Finally it presents the "Minimal Cognitive Grid", a pragmatic methodological tool proposed to rank the different degrees of structural accuracy of artificial systems in order project and predict their explanatory power.

Chapter 4 – **Examples of Cognitively Inspired Systems and application of the MCD**

Given the proposal presented in the previous sections of the book, this chapter describes some practical applications of the Minimal Cognitive Grid by showing how it allows to collocate different types of artificial systems in the landscape formed by the cognitive design approach.

Examples of artificial models of cognition and cognitive architectures will be shown and compared with examples of functionalists AI systems that, despite called as instances of "cognitive computing", cannot be considered realistic models of our cognition.

Chapter 5 - **Evaluating the Performances of Artificial Systems**

This chapter introduces the main proposals that have been developed in oder to evaluate the performance of artificial systems (cognitively inspired or not) and to justify the ascription of faculties coming from the "cognitive" vocabulary (like "intelligence") to such systems. After introducing the Turing Test, its problematic aspects and some of the main modifications proposed (e.g. the Super Turing Test and other variations), we will analyze other frameworks like the Newell Test for a theory of cognition and other tasks and challenges that have been used - with different purposes - as a testbed for the evaluation of artificial systems. These tasks go from the Robocup World Soccer to the DARPA Challenges for autonomous vehicles to the recently proposed Winograd Schema Challenge and the RoboCup@Home. We will analyse these proposal both in the light of their eventual explanatory role in the context of a computationally-driven science of the mind and with respect to their actual capacity of evaluating the "intelligence" of artificial systems.

Chapter 6 - The Next Steps

This concluding chapter will synthesize the main issues presented along the book and will try to provide a roadmap for the next years in the context of cognitive AI research, by suggesting fields where the cognitive design approach can provide valuable inputs for the realization of better AI systems.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. Psychological review, 111(4), 1036.

Anderson, J. R., & Lebiere, C. (2003). The Newell test for a theory of cognition. *Behavioral and brain Sciences*, *26*(5), 587-601.

Anderson, J.R., Betz, J., (2001) A hybrid model of categorization, Psychonomic Bulletin & Review 8 (4) (629–647).

Anderson, J. R. (1990). The adaptive character of thought . New York: Psychology Press.

Arbib, M. A. (2018). From cybernetics to brain theory, and more: A memoir. *Cognitive Systems Research*, *50*, 83-145.

Arbib, M. A. (ed) (2002). The Handbook of Brain Theory and Neural Networks.MIT Press.

Augello, A., Infantino, I., Lieto, A., Pilato, G., Rizzo, R., & Vella, F. (2016). Artwork creation by a cognitive architecture integrating computational creativity and dual process approaches. *Biologically inspired cognitive architectures*, *15*, 74-86.

Ball, J., Rodgers, S., & Gluck, K. (2004). Integrating ACT-R and Cyc in a large-scale model of language comprehension for use in intelligent agents. In *AAAI workshop* (pp. 19-25).

Bartlett, F. (1958). Thinking: An experimental and social study. Allen and Unwin.

Bermudez, J.L. 2005. *Philosophy of Psychology*, Routledge, London and New York.

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk, Econometrica, Vol. 22, No. 1, pp. 23-36

Blouw, E. Solodkin, P. Thagard, C. Eliasmith, Concepts as semantic pointers: a framework and computational model, Cognitive science.

Boden, M. A. (2009). Computer models of creativity. *AI Magazine*, *30*(3), 23-23.

Bonatti, P. A., Faella, M., Petrova, I. M., & Sauro, L. (2015). A new semantics for overriding in description logics. *Artificial Intelligence*, *222*, 1-48.

Bossaerts, P., Yadav, N., & Murawski, C. (2019). Uncertainty and computational complexity. Philosophical Transactions of the Royal Society B: Biological Sciences , 374 (1766), 20180138.

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. MIT press.

Bringsjord, S. (2011). Psychometric artificial intelligence.Journal of Experimental & Theoretical Artificial Intelligence,23(3), 271–277.

Breazeal, C. L. (2004). *Designing sociable robots*. MIT press.

Breazeal, C. L and Scassellati B. (2000). Infant-like social interactions between a robot and a human caretaker. Adaptive Behavior, 8(1).

Brooks, R. A. (1999). *Cambrian intelligence: The early history of the new AI*. MIT press.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, *47*(1-3), 139-159.

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE journal on robotics and automation*, *2*(1), 14-23.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Agarwal, S. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Cangelosi, A., & Parisi, D. (Eds.). (2012). *Simulating the evolution of language*. Springer Science & Business Media.

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model* (Vol. 18). John Wiley & Sons.

-Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. Organizational Behavior and Human Decision Processes , 90 (1), 63-86.

Chater, N. & Oaksford, M. (1999). Ten years of the rational analysis of cognition. Trends in cognitive sciences , 3 (2), 57-65.

Chiodino, E., Di Luccio, D., Lieto, A., Messina, A., Pozzato, G. L., & Rubinetti, D. (2020). A Knowledge-based System for the Dynamic Generation and Classification of Novel Contents in Multimedia Broadcasting. In *Proceedings of ECAI* 2020, p. 680-687

Chiodino, E., Lieto, A., Perrone, F., & Pozzato, G. L. (2020). A goal-oriented framework for knowledge invention and creative problem solving in cognitive architectures. *Proceedings of ECAI 2020*, p. 2893-2894.

Chomsky, N. (1957). *Syntactic structures*. Mouton & Co.

Churchland, P.S. and Sejnowski, T.J. (1992), The Computational Brain, MIT Press, Cambridge, Mass.

Churchland, P.M. and Churchland P.S. (1990), Could a machine think? Recent arguments and new prospects, Scientific American, 262 (1): 32-37.

Copeland B.J. (2000). The Turing test. Minds and Machines, 10: 519-539.

Cordeschi, R., & Frixione, M. (2007). Computationalism under attack. In *Cartographies of the Mind* (pp. 37-49). Springer, Dordrecht.

Cordeschi R. (2002). The Discovery of the Artificial. Behavior, Mind and Machines Before and Beyond Cybernetics. Berlin and New York: Springer. Doi: 10.1007/978-94-015-9870-5

Cordeschi, R. (1991). The discovery of the artificial. some protocybernetic developments, 1930–1940. AI & society, 5 , 218–238.

Crick, F. (1989) The recent excitement about neural networks. Nature 337:129–32.

Damasio, A. (1994). Descartes' Error: Emotion, Reason, and the Human Brain Paperback.

Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, *58*(9), 92-103.

de Melo, C. M., & Terada, K. (2019). Cooperation with autonomous machines through culture and emotion. *PloS one*, *14*(11), e0224758.

Dennett, D. C. (1988). The Intentional Stance. Cambridge, MA: Bradford Books: MIT Press.

Dennett, D. C. (1986). Is there an autonomous 'knowledge level'? In Pylyshyn, Z. W., & Demopoulos, W. (Eds.), Meaning and Cognitive Structure: Issues in the computational theory of mind. Norwood, NJ: Ablex.

Dennett, D. C. (1976). Brainstorms: Philosophical essays on mind and psychology. Montgomery, VT: Bradford Books.

Derbinsky, N., Laird, J. E., & Smith, B. (2010). Towards efficiently supporting large symbolic declarative memories. In *Proceedings of the 10th international conference on cognitive modeling* (pp. 49-54).

Dreyfus, H. (1972). What computers can't do: The limits of artificial intelligence. MIT Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *Science*, 338(6111), 1202-1205.

Epstein, R., Roberts, G., & Beber, G. (Eds.). (2009). *Parsing the Turing test* (pp. 978-1). Springer Netherlands.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. Psychological review, 87(3), 215.

Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: beyond jeopardy!. *Artificial Intelligence*, *199*, 93-105.

Fikes, R. E.; Hart, P. E.; and Nilsson, N. J. (1972). Learning and Executing Generalized Robot Plans. Artificial Intelligence 3(4): 251–288.

Fodor, J.A. and Pylyshyn, Z.W. (1988), Connectionism and cognitive architecture: a critical analysis, Cognition, 28: 3-71

Fodor, J.A. (1986), Why Paramecia don't have mental representations, Midwest Studies in Philosophy, 10: 3-23.

Forbes, N. (2004). *Imitation of life: how biology is inspiring computing*. Cambridge MA: Mit Press.

Forbus, K. D., & Hinrich, T. (2017). Analogy and relational representations in the companion cognitive architecture. *AI Magazine*, *38*(4), 34-42.

Franklin, S. (1995). Artificial minds. MIT Press.

French R. (1990). Subcognition and the limits of the Turing Test, Mind, 99: 53-65.

Friedman, D., Isaac, R. M., James, D., & Sunder, S. (2014). Risky curves. On the empirical failure of expected utility. New York: Routledge.

Frixione, M. (2015). The Turing Test and the interface problem: a role for the imitation game in the methodology of cognitive science. *PARADIGMI*.

Frixione, M. (2001). Tractable competence. *Minds and Machines*, *11*(3), 379-397.

Gärdenfors, P. (2000). Conceptual spaces: The geometry of thought. Cambridge, MA: MIT Press.

Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences*. Hachette Uk.

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy. Cognitive Science 7(2): 155–170. doi.org/10.1207/ s15516709cog0702_3

Gerndt, R., Seifert, D., Baltes, J. H., Sadeghnejad, S., & Behnke, S. (2015). Humanoid robots in soccer: Robots versus humans in RoboCup 2050. *IEEE Robotics & Automation Magazine*, *22*(3), 147-154.

Gigerenzer, G. (2019). How to explain behavior? Topics in Cognitive Science.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). Heuristics: The foundations of adaptive behavior . New York: Oxford Universtity Press (p. xx).

Gigerenzer, G. (2000). Adaptive thinking: Rationality in the real world. Oxford: Oxford University Press.

Gigerenzer, G., & Todd, P. M. (1999). Simple heuristics that make us smart. Oxford University Press, USA.

Giordano, L., Gliozzi, V., Lieto, A., Olivetti, N., & Pozzato, G. L. (2020). Reasoning about Typicality and Probabilities in Preferential Description Logics. In Applicayions and Practices in Ontology Design, Extraction and Reasoning, Studies on Semantic Web. IOS Press. *arXiv preprint arXiv:2004.09507*.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). Cambridge: MIT press.

Goldstein, D. G. & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. Psychological Review , 109 (1), 75.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. Topics in cognitive science , 7 (2), 217-229.

Grossberg, S. (1982), A psychophysiological theory of reinforcement, drive, motivation, and attention, Journal of Theoretical Neurobiology, l: 286-36.

Harnad S. (2001). Minds, machines and Turing. Journal of Logic, Language and Information, 9: 425-445. Doi: 10.1023/A:100831530886.

Hamad, S. (1990), The symbol grounding problem, Physica D, 42: 335-346.

Hart, P. E.; Nilsson, N. J.; and Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths in Graphs, IEEE Transactions on Systems Science and Cybernetics SSC-4(2): 100–107.

Hawkins, J., and Blakeslee, S. 2005. On Intelligence. New York: Times Books.

Hebb, D.O. (1949), The Organization of Behavior, Wiley and Chapman, New York and London.

Hempel, C. and P. Oppenheim., (1948) 'Studies in the Logic of Explanation.', *Philosophy of Science*, 15: 135–175. Reprinted in Hempel, 245–290, 1965°

Hofstadter, D. R., & Mitchell, M. (1995). The copycat project: A model of mental fluidity and analogy-making. *Advances in connectionist and neural computation theory*, 2, 205-267.

Holland. J. (1975), Adaptation in natural and artificial systems . University of Michigan Press.

Hopfield, J.J. (1982), Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences of USA, 79: 2554-2558

Horvitz, E. J., Cooper, G. F., & Heckerman, D. E. (1989). Reflection and action under scarce resources: Theoretical principles and empirical study. *IJCAI*, 2, 1121–1127.

Hsu, Feng-Hsiung (1999), "IBM's Deep Blue Chess Grandmaster Chips", IEEE Micro, Los Alamitos, CA, USA: IEEE Computer Society, 19 (2): 70–81.

Jean, S., Cho, K., Memisevic, R., & Bengio, Y. (2015). On using very large target vocabulary for neural machine translation. *Proc. ACL-IJCNLP, arXiv preprint arXiv:1412.2007*.

Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness (Vol. 6). Harvard University Press.

Johnson-Laird, P. N. (2006). How we reason. USA: Oxford University Press.
Kahneman, D. 2011. Thinking, Fast and Slow. London: Allen Lane

Kakas, A. C., & Michael, L. (2016). Cognitive Systems: Argument and Cognition. *IEEE Intell. Informatics Bull.*, *17*(1), 14-20.

Kitano, H., & Asada, M. (2000). The RoboCup humanoid challenge as the millennium challenge for advanced robotics. *Advanced Robotics*, *13*(8), 723-736.

Kitano, H., Hamahashi, S. and Luke, S. (1998), The perfect C. elegans project: an initial report, Artificial Life, 4: 141-156.

Kotseruba, Iuliia, and John K. Tsotsos.(2020). 40 years of cognitive architectures: core cognitive abilities and practical applications." *Artificial Intelligence Review* 53.1: 17-94.

LaCurts, K. (2011). Criticisms of the turing test and why you should ignore (most of) them. CSAIL, MIT *Philosophy and theoretical computer science*.

Laird, J. E., Gluck, K., Anderson, J., Forbus, K. D., Jenkins, O. C., Lebiere, C., ... & Wray, R. E. (2017). Interactive task learning. *IEEE Intelligent Systems*, *32*(4), 6-21.

Laird, J.; Lebiere, C.; and Rosenbloom, P. 2017. A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. AI Magazine 38(4). doi.org/10.1609/aimag. v38i4.2744.

Laird, J. 2012. The SOAR Cognitive Architecture. Cambridge, MA: The MIT Press

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. Artificial Intelligence, 33, 1–64.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*.

Langley, P. (2019). Scientific discovery, causal explanation, and process model induction. *Mind & Society*, *18*, 43-56.

Langley, P. (2017). Interactive cognitive systems and social intelligence. IEEE Intelligent Systems, 32(4), 22-30.

Langley, P. (2012). The cognitive systems paradigm. Advances in Cognitive Systems, 1, 3–13.

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*(2), 141-160.

Langley, P., Bradshaw, G. L., & Simon, H. A. (1983). Rediscovering chemistry with the BACON system. In *Machine learning* (pp. 307-329). Springer, Berlin, Heidelberg.

Lenat, D.B, (1995) Cyc: A large-scale investment in knowledge infrastructure, Communications of the ACM 38 (11), 33–38.

Levesque, H. J. (2017). *Common sense, the Turing test, and the quest for real AI*. MIT Press.

Levesque, H., Davis, E., & Morgenstern, L. (2012). The winograd schema challenge. In *Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*.

Lieder, F. & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences , 1-85.

Lieto, A., & Pozzato, G. L. (2020). A description logic framework for commonsense conceptual combination integrating typicality, probabilities and cognitive heuristics. *Journal of Experimental & Theoretical Artificial Intelligence*, Vo. 32 (5), 769-804.

Lieto, A. (2020). Bounded Rationality and Heuristics in Humans and in Artificial Cognitive Systems. *Isonomia*, *2037*, 53-73.

Lieto, A. (2019). Heterogeneous proxytypes extended: Integrating theory-like representations and mechanisms with prototypes and exemplars. In *Biologically Inspired Cognitive Architectures Meeting* (pp. 217-227). Springer, Cham.

Lieto, A., Perrone, F., Pozzato, G. L., & Chiodino, E. (2019). Beyond subgoaling: A dynamic knowledge generation framework for creative problem solving in cognitive architectures. *Cognitive Systems Research*, *58*, 305-316.

Lieto, A., Lebiere, C., & Oltramari, A. (2018). The knowledge level in cognitive architectures: Current limitations and possible developments. *Cognitive Systems Research*, *48*, 39-55.

Lieto, A., Bhatt, M., Oltramari, A., & Vernon, D. (2018). The role of cognitive architectures in general artificial intelligence. . *Cognitive Systems Research*, *48*, 1-3.

Lieto, A., Radicioni, D. P., & Rho, V. (2017). Dual PECCS: a cognitive system for conceptual representation and categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*(2), 433-452.

Lieto, A., Radicioni, D., Rho, V., & Mensa, E. (2017b). Towards a unifying framework for conceptual represention and reasoning in cognitive systems. *Intelligenza Artificiale*, *11*(2), 139-153.

Lieto, A. (2017). Representational limits in cognitive architectures. In *EUCognition Meeting (European Society for Cognitive Systems)" Cognitive Robot Architectures"* (Vol. 1855, pp. 16-20). Ceur-ws.

Lieto, A., Radicioni, D. P., & Rho, V. (2015). A common-sense conceptual categorization system integrating heterogeneous proxytypes and the dual process of reasoning. In *24th International Joint Conference on Artificial Intelligence (IJCAI 2015)* (pp. 875-881). AAAI Press.

Lieto, A. (2014). A computational framework for concept representation in cognitive systems and architectures: Concepts as heterogeneous proxytypes, Procedia Computer Science, 41, 6-14.

Malt, B. (1989) An on-line investigation of prototype and exemplar strategies in classification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15 (4), 539.

Marr, D. (1977). "Artificial intelligence - A personal view," *Artificial Intelligence* 9: 37-48.

Marr, D. (1982). *Vision*. San Francisco : W . H . Freeman.

McCarthy, J. (2007). From here to human-level ai. Artificial Intelligence, 171, 1174–1182.

McCarthy, J. (1980). Circumscription—a form of non-monotonic reasoning. *Artificial intelligence*, *13*(1-2), 27-39.

McCarthy, J., and Hayes, P. (1969). Some Philosophical Problems from the Standpoint of Artificial Intelligence. In Machine Intelligence 4, ed. B. Meltzer and D. Michie. Edinburgh, Scotland: Edinburgh University Press.

McCarthy, J. (1960). *Programs with common sense* (pp. 300-307). RLE and MIT computation center.

McCarthy, J.; Minsky, M. L.; Rochester, N.; and Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. AI Magazine 27(4): 12–14.

McClelland, J. L. (2010). Emergence in cognitive science. Topics in cognitive science, 2(4), 751-770.

McCulloch, W.S. and Pitts, W. (1943), A logical calculus of the ideas immanent in nervous activity, Bulletin of Mathematical Biophysics, 5: 115-137. Reprinted in J.A. Anderson and E. Rosenfeld (1988).

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., & Bernardino, A. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural networks*, *23*(8-9), 1125-1134.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38.

Miller, G. A. (1956), The magical number seven. Psychological Review. 63, 81.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, *38*(11), 39-41.

Minkowski, M. (2013). Explaining the Computational Mind. MIT Press.

Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster.

Minsky, Marvin (1986). The Society of Mind. New York: Simon & Schuster.

Minsky, M. (1975). A framework for representing knowledge, in: The Psychology of Computer Vision , Patrick H. Winston (ed.), McGraw-Hill, New York.

Minsky, M.L. and Papert, S. (1969), Perceptrons, MIT Press, Cambridge, Mass

Morgenstern, O., & Von Neumann, J. (1953). *Theory of games and economic behavior*. Princeton university press.

Murphy, G. (2002). *The big book of concepts*. MIT press.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, MA: Harvard University Press.

Newell, A. (1988). The intentional stance and the knowledge level. *Behavioral and Brain Sciences*, *11*(3), 520-522.

Newell, A. The knowledge level. Artificial Intelligence, 1982,75,87-127.

Newell, A. (1980). Physical symbol systems. Cognitive Science 4:135–83.

Newell A., Simon, H.A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126.

Newell, A. (1973). You can't play a game of 20 questions with nature and win. *Visual information processing*, 238-308.

Newell, A., and Simon, H. A. (1972) Human problem solving. Englewood Cliffs, NJ: Prentice-Hall

Newell, A., Shaw, J.C . and Simon, H. A. (1958), Elements of a theory of human problem-solving, Psychological Review, 65: 151-166

Nilsson, N. J. Problem solving methods in artificial intelligence. New York: McGraw-Hill, 1971.

Ohlsson S, R. H. Sloan, G. Tur´an, D. Uber, A. Urasky, (2012). An approach to evaluate ai commonsense reasoning systems., in: FLAIRS Conference.

Pearl, J., & Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic Books.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Picard, R. W. (1997). *Affective computing*. MIT press.

Piccinini, G. (2007). Computational modelling vs. Computational explanation: Is everything a Turing Machine, and does it matter to the philosophy of mind?. Australasian Journal of Philosophy, 85(1), 93-115.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1-2), 73-193.

Putnam, H. (1960). Minds and machines.

Pylyshyn, Z.W. (1989), "Computing in Cognitive Science", pp.51-91 in Posner, M.I.(ed.), *Foundations of Cognitive Science*, The MIT Press, (Cambridge).

Pylyshyn, Z. W. (1984), Computation and Cognition. Toward a Foundation for Cognitive Science, MIT Press, Cambridge, Mass.

Pylyshyn, Z.W. (1979), Complexity and the study of artificial and human intelligence, in M. Ringle (ed.), Philosophical Perspectives in Artificial Intelligence, Harvester, Brighton.

Quillian, M.R. (1968). Semantic memory, in M. Minsky (ed.), *Semantic Information Processing*, MIT Press, Cambridge, MA.

Reiter, R. (1980). A logic for default reasoning. Artificial intelligence, 13(1-2), 81-132.

Rich, P., Blokpoel, M., de Haan, R., & van Rooij, I. (2020). How intractability spans the cognitive and evolutionary levels of explanation. Topics in Cognitive Science.

Rosch, E. (1975): Cognitive representations of semantic categories. J. Exp. Psychol. Gen. 104(3), 192–233.

Rosenblatt, F. (1958), The Perceptron: a probabilistic model for information storage and organization in the brain, Psychological Review, 65: 386-408.

Rosenbloom, P. S. (2013). The Sigma cognitive architecture and system. *AISB Quarterly*, *136*, 4-13.

Rosenblueth, A. and Wiener, N. (1945), The role of models in science, Philosophy of Science , 12: 316- 321.

Rosenblueth, A., Wiener, N. and Bigelow, J. (1943), Behavior, purpose and teleology, Philosophy of Science, 10: 18-24.

Rumelhart, D.E., McClelland, J.L. and the PDP Research Group (1986), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 2 vols., MIT Press, Cambridge, Mass.

Russell S.J., Norvig P. (2002) *Artificial Intelligence. A Modern Approach (2nd Ed.)* Prentice Hall, Englewood. Cliffs, NJ.

Russell, S. J., & Subramanian, D. (1995). Provably bounded-optimal agents. *Journal of Artiicial Intelligence Research*, *2*, 575–609.

Salvucci, D. (2014). Endowing a cognitive architecture with world knowledge. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Samuel, A.L. (1959), Machine learning, The Technology Review, 62 (1): 42-45.

Searle, J., (1999), 'The Chinese Room', in R.A. Wilson and F. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*, Cambridge, MA: MIT Press.

Searle, J., (1980), 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: 417–57.

Schank, R.C. and Abelson, R.P. (1977). Scripts, Plans, Goals and Understanding, Erlbaum, Hillsdale, N.J.

Schank, R.C. and Nash-Webber, B.L. (eds.) (1975), Theoretical Issues in Natural Language Processing, Bolt, Beranek and Newman, Cambridge, Mass.

Schank, R. C. (1972). Conceptual dependency: A theory of natural language understanding. Cognitive Psychology, 3 , 552–631.

Shayani, H. (2013). A Practical Investigation into Achieving Bio-Plausibility in Evo-Devo Neural Microcircuits Feasible in an FPGA. PhD Thesis. UCL.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Chen, Y. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354-359.

Simon, H.A. (1981). The Sciences of the Artificial, MIT Press, Cambridge, Mass., 2nd edition: MIT Press, Cambridge, Mass.

Simon, H. A. (1979). *Models of thought* (Vol. 352). Yale university press.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99-118.

Simon, H. A. (1947). Administrative behavior: A study of decision-making processes in administrative organization). New York: Macmillan.

Smith, J.D., Minda, J.P. (1998): Prototypes in the mist: The early epochs of category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition 24(6), 1411.

Smith, J. D., Murray, M. J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 23, 659-68.

Sloman, A. (2014). How can we reduce the gulf between artificial and natural intelligence?. In *AIC* 2014, 2nd International Workshop on Artificial Intelligence and Cognition, Turin, pp. 1-13.

Sloman A (2002) How many separately evolved emotional beasties live within us? In: Trappl R, Petta P, Payr S (eds) Emotions in Humans and Artifacts, MIT Press, Cambridge, MA, pp 29–96.

Sloman A (2001) Beyond Shallow Models of Emotion. Cognitive Processing: International Quarterly of Cognitive Science 2(1):177–198.

Sloman, A. (1998). Damasio, Descartes, alarms and meta-management. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)* (Vol. 3, pp. 2652-2657). IEEE.

Sun, R. (2007). The importance of cognitive architectures: An analysis based on CLARION. *Journal of Experimental & Theoretical Artificial Intelligence*, *19*(2), 159-193.

Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, *17*(3), 341-373.

Susanto Y., A Livingstone, BC Ng, E Cambria. The Hourglass model revisited. IEEE Intelligent Systems 35(5) (2020)

Turing A. M. (1950). Computing machinery and intelligence. Mind, 59(236):433–460.

Turing, A.M. (1936-37), On computable numbers, with an application to the Entscheidungsproblem, Proceeding of the London Mathematical Society (2''' series), 42: 230-265; 43: 544.

Tversky A. and Kahneman D. (1983). Extension versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90,4: 293-315.

Van Fraassen, B. C. (1980).The scientific image.Oxford: Clarendon Press

Veale, T., & Cardoso, F. A. (Eds.). (2019). *Computational creativity: The philosophy and engineering of autonomously creative systems*. Springer.)

Vernon, D., von Hofsten, C., & Fadiga, L. (2017). Desiderata for developmental cognitive architectures. *Biologically Inspired Cognitive Architectures*, *18*, 116-127.

Vernon, D.,(2014) Artificial cognitive systems — A primer, MIT Press.

Wallin, A., & Gärdenfors, P. (2000). Smart people who make simple heuristics work. Behavioral and Brain Sciences, 23(5), 765-765.

Watson, J.B. (1913), Psychology as the behaviorist views it, Psychological Review, 20: 158-177.

Webb, B. (2001). Can robots make good models of biological behaviour?. *Behavioral and brain sciences*, *24*(6), 1033-1050.

Weizenbaum, J. (1976), Computer Power and Human Reason, Freeman: San Francisco.

Wiener, N. (1948/1961), Cybernetics, or Control and Communication in the Animal and the Machine, MIT Press , Cambridge, Mass. (2''' edition: MIT Press, Cambridge, Mass., 1961).

Winograd, T. (1972), Understanding Natural Language. Academic Press: New York.

Winston, P. H. (2014). *The genesis story understanding and story telling system a 21st century step toward artificial intelligence*. Center for Brains, Minds and Machines (CBMM).

Winston, P. H. (2012a). The next 50 years: a personal view. *Biologically Inspired Cognitive Architectures*, *1*, 92-99.

Winston, P. H. (2012b). The right way. Advances in Cognitive Systems, 1:23–36, 2012.

Winston, P. H. (2011). The strong story hypothesis and the directed perception hypothesis. In Pat Langley, editor, Technical Report FS-11-01, Papers from the AAAI Fall Symposium, pages 345–352, Menlo Park, CA, 2011. AAAI Press.

Zaadnoordijk, L., Besold, T. R., & Cusack, R. (2020). The Next Big Thing (s) in Unsupervised Machine Learning: Five Lessons from Infant Learning. *arXiv preprint arXiv:2009.08497*.

Zadeh, L. A. (1988). Fuzzy logic. *Computer*, *21*(4), 83-93.