A Standard Model of Mind:
AAAI Technical Report FS-17-05

Towards a Standard Cognitive Framework for Socially Oriented, Adaptive, and Generative Human-Environment Agents

Jens K. Madsen, Richard Bailey, Ernesto Carrella

School of Geography and the Environment, University of Oxford, South Parks Road, Oxford
jens.madsen@ouce.ox.ac.uk // richard.bailey@ouce.ox.ac.uk // ernesto.carrella@ouce.ox.ac.uk

Toby Pilditch

Department of Experimental Psychology, University College London, Malet Street, London
t.pilditch@ucl.ac.uk

Abstract

While several unified theories of cognition have been proposed, no framework has been established with the same degree of universal agreement as in biology and physics. A universal model of cognition is needed to direct research, push cognitive sciences, and test more or less realistic interventions on shifting environments. Here, we propose the necessary components for modelling a socially oriented, generative, and adaptive agent. We argue such a model requires modules for information input, management, storage, and use in order to grow an agent capable of human-like adaptive, socio-cultural behavioural strategies. We further argue that such an agent may be tested in different contexts through Agent-Based Modelling.

Introduction

Standard models offer scientific disciplines a unified theoretical perspective that may account for existing data sets across observations, generate predictions of unknown aspects and provide a theoretical approach that is empirically falsifiable. In biology, the Theory of Evolution offers a theoretical framework for continuous speciation through anagenesis and extinction against the backdrop of natural selection. In physics, the Standard Model, though currently incomplete, offers a unified theory of particle interaction. These models offer frameworks for biology and physics.

Centrally, both the Theory of Evolution and the Standard Model in physics are *predictive* and *generative* models that are empirically falsifiable. For example, the standard model required the existence of the Higgs Boson (or something functionally equivalent). Had CERN failed to detect the Higgs Boson, the Standard Model would have required substantial revision or, potentially, wholesale rejection. Similarly, the Theory of Evolution can make falsifiable predictions. Our simian cousins have 24 chromosome pairs just like our common ancestors while humans only have 23 pairs. Consequently, the theory of evolution predicts that humans have a fused chromosome. Chromosome #2 was found to have this exact property. If this chromosome fusion could not have been detected, the Theory of Evolution would require considerable revision. Crucially, the Theory of Evolution allows for different species to emerge *from the same principles* if contextual pressures change. That is, if the environment and experienced life of the organism differs between two places, the same generative evolutionary process will yield different organisms over time. For example, turtles on different Galapagos Islands have evolved different neck lengths as an adaptive feature.

Although cognitive sciences have been prevalent and growing in theoretical scope since the 1950s, no comparative unified framework has hitherto been established with the same degree of collective agreement as the standard frameworks in biology and physics – though several prominent models such as ACT-R (Anderson, 1983), SOAR (Laird, 2012), and Agent_Zero (Epstein, 2013) have been proposed. This may not be unsurprising, as humans are difficult to predict and formally describe. Comparatively, electrons do not have socio-cultural preferences or sampled memories of previous interactions with other particles. As a consequence of this, a universally agreed upon standard model is extremely difficult to develop.

Accordingly, most studies in cognitive science has collected data on and computationally modelled specific issues (descriptive or predictive) rather than a generative framework that can be implemented across situations. Branching across biology, the current observed taxonomy is akin to early collection of biological classification of different species before a generative theory of evolution was proposed. Rather than a taxonomy of computationally expressible descriptive functions, a standard model of cognition must be process-oriented, generative, and provide falsifiable predictions. Ideally, a framework will identify key cognitive components related to cognition in a formally expressible way that is testable beyond the confinements

of a particular task. Further, the framework must allow for the development of socio-cultural contracts and adaptive behavioural strategies.

Insights from agent-based modelling (e.g. Epstein & Axtell, 1997; Miller & Page, 2007; Gilbert, 2008) further demonstrate the necessity for the generation of a standard model of cognition. While cognitive computational models typically take point of departure in computational functions that are analytically solvable, many cognitive tasks occur in complex environments that cannot be resolved analytically. For example, an analytic model cannot prescribe the aggregate investment strategies of a heterogeneous investor population if they can interact with each other over time.

If cognitive scientists are to implement cognitively realistic agents in agent-based models, the need for a standard model of cognition that goes beyond computational or explanatory modelling of a particular cognitive phenomenon or task increases. In particular, if researchers aim to develop models of human decision-making, reasoning or other crucial cognitive functions that go beyond custom-made solutions to specific issues, a standard model is necessary to test functionality across issues ranging over policy-making, intervention testing, and predictions of emergent behavioural patterns.

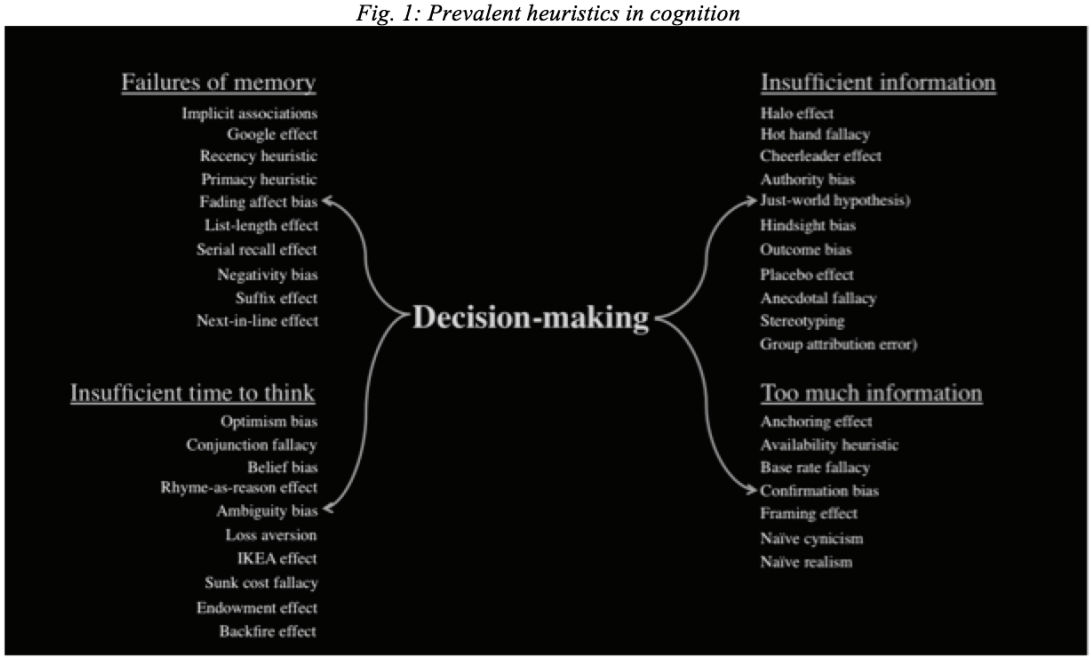
In this position paper, we propose the foundational elements we believe must be integrated in a standard model of cognition for socially oriented, adaptive, and generative agents. We suggest that a standard model of cognition must include formally expressible models of information acquisition and input, information management, memory, and information use, as well as modifying components that allow for individual differences. This is in line with SOARs modular composition. Importantly, we do not propose specific module solutions, as the intention is not to argue for one belief revision model over another (e.g. Bayesian versus frequentist updating). Rather, the paper proposes the necessary elements needed to develop a socially oriented, adaptive, and generative agent and thus a standard framework for human-like cognition.

Going beyond a heuristic taxonomy

With notable exceptions (e.g. Kemp et al., 2005), cognitive computational models have predominantly focussed on task-specific issues such as integration of new information in reasoning and argumentation (Hahn & Oaksford, 2007), argument structures (Harris et al., 2012), perception of information from other sources (Harris et al., 2015), memory recall (Lange et al., 2013), information sampling (Stewart et al., 2006), information search strategies (Wu et al., 2017), heuristic decision-making strategies (Kahneman, 2011), sampling strategies (Stewart et al., 2006), and reasoning strategies (Koralus & Mascarenhas, 2013). Each of these issues point to essential components of cognition, as they relate to how information can be acquired, manipulated, and used in specific contexts. Specifically, the models developed to describe or potentially predict how humans perform these tasks are essential to appreciate cognition more generally. To illustrate the issue-oriented modelling approach, figure 1 summarises key cognitive functions identified in decision-making.

The computational models of specific cognitive tasks are critical in going towards a standard model of cognition. Equally, the observations in reasoning and decision-making studies that show modifying influences are crucial, as they provide an empirical benchmark against which standard models of cognition can be measured (as a standard model should be able to account for the observed data).

Fig. 1: Prevalent heuristics in cognition



However, an issue-specific, non-generative approach falls short of a standard model of cognition for several reasons. First, it is insufficient to model (descriptively or predictively) cognitive functions such as information integration computationally in isolation from the function of the overall cognitive apparatus. Much like

neutrinos are a key, but insufficient element of the standard model in physics, isolated cognitive tasks are peaking under a small patch of the blanket of a standard model of cognition. Second, the dual-process account does not specify a process through which specific heuristics are chosen. If a standard model of cognition is to be achieved, it is necessary for the *same* process to be able to yield different cognitive or behavioural outcomes if parameters or the situation changes. Third, as discussed in the following, most human activity occurs in complex situations where isolated functions are unsatisfactory. In particular, the interaction between key elements (e.g. memory, attention, and information integration) requires a standard model of cognition that can approach novel and multifaceted tasks. This should yield falsifiable predictions for testing proposed standard models of cognition. While many cognitive studies deal with issue-specific heuristic strategies, important proposals toward a unified theory of cognition have been made.

Unified cognitive models

The fields of artificial intelligence and cognitive sciences have moved towards a general framework for cognition. Three of the most prominent unified cognitive architectures in the fields are ACT-R, SOAR, and Agent_Zero. These describe the computational components necessary for general agents, i.e. agents who can perform central cognitive capabilities found in humans such as decision-making, learning, and natural language understanding.

ACT-R describes perceptual-motor and memory modules. The former places the agent in the world while the latter allows for declarative and procedural memory states. SOAR allows for bottom-up, parallel, and automated processes, deliberation of belief-states, and procedural rules for solving sub-states given uncertainty. This builds on the Problem Space hypothesis, universal sub-goaling, a modular composition, and symbolic and relational memory. The modules are concerned with decision-making, memory, learning, and with perceptual-motor functions. The framework proposed in this paper for a socially oriented, generative, and adaptive (potentially heterogeneous) agent is also highly modular, as it describes the necessary components needed for social functions and practices to occur. Finally, Agent_Zero uses procedural rules for information management and use as well as the emotional state of the agent.

The above-mentioned unified theories are important steps toward a standard model. A general framework for cognition should be able to manifest at least the following properties. First, the agent should be adaptive to novel environments and problems. That is, the same cognitive architecture should yield sensible decisions when fishing (such as fishing the line), protesting (such as rioting, Epstein, 2013), trading (such as loss aversion), and poaching (such as learning migration patterns of the prey). These behavioural patterns should replicate observable data such as fishing the line or heuristic or biased strategies without hard-wiring these into the agent. Second, the agent should be socially engaged and able to communicate, cheat, and make use of the social structure to inform decisions. Social structures and practices should emerge *from* interaction between agents rather than given *to* the agents. For example, given contextual incentives, agents should learn to cooperate and defect in manners that replicate observed strategies. Third, the agent should be able to learn and change strategies within dynamic environments. Finally, a general framework for cognition should yield testable predictions (e.g. speed of behavioural adaption to novel policies in fisheries). These should include moderate behavioural outcomes (such as sensible fishing strategies) and extreme behavioural outcomes (such as becoming an extremist). Crucially, the general framework should yield these different behavioural adaptations without changing the cognitive components. That is, the social (desirable as well as undesirable) structures and practices should emerge over time as agents adapt to the environment and other agents when given a goal and preferences.

We believe the social components and preference ordering, crucial as they are to human behaviour, have not been specified sufficiently within previous unified theories.

In short, a general cognitive framework should be implementable across dynamic environments and be able to solve a range of goal-oriented tasks, which then yield emergent behavioural patterns, individually and socially. From this, behavioural and social strategies should emerge. Pattern-oriented simulations may be used to test the validity of the model or predictively to test the supposed outcome of a particular set of interventions.

In the paper, we present the minimal elements needed to develop a (potentially) heterogeneous, ***socially oriented***, ***generative*** goal-oriented agent who can ***adapt*** to dynamic environments.

Towards a socially oriented, adaptive, and generative cognitive framework

The Theory of Evolution relies on several broad mechanisms such as responsiveness to external factors, organism-oriented goals (e.g. acquire food or reproduce), and DNA blending upon procreation. Each of these broader mechanisms necessarily consists of several sub-mechanisms. From these mechanisms, the standard model of evolution allows for development of species variation, changes, and different adaptive strategies. Important work suggests the usefulness of generative models of cognition (Stuhlmüller et al., 2010; Tenenbaum et al., 2011). To expand a generative cognitive perspective more generally, we consider the broad mechanisms that must *necessarily* be involved in a standard model of cognition, which is *socially oriented*, *generative*, *adaptive* (i.e. an agent who can make use of its

social capital and can adapt to different problems and contexts to pursue a goal-state).

Information input A standard model of cognition must necessarily provide adequate expression of how the organism searches for and inputs new information. Further subdividing the input mechanism, a standard model of cognition must be able to describe the functional motivation for addressing the present problem. That is, the agent must have an underlying motivated goal, ranging from acquisition of better information (e.g. learning about the environment), to acquiring money, or finding a match on a dating website. As several beliefs can be used to solve a problem (e.g. ethical considerations, earning potential, social standing, etc.), the agent must have a preference ordering. Second, an agent must have a perceptual-motor module for perceiving the environment. Further, agents must be able to search strategically for new information. Such information may be gained from sampling the world (direct evidence) or from consulting other agents through the social network (indirect evidence). Related to this, the agent must have a way to direct attention to the information that the search strategies for the goal-state yield. Further, as agents cannot cope with all available information at the same time, the model will require a search scope.

Information management Having encountered, engaged with, and encoded the information deemed relevant for the present goal-state, the agent needs to be able to manipulate the information and integrate it within pre-existing epistemic frameworks. The agent must be capable of updating beliefs given direct evidence (e.g. from sampling from the world). Additionally, as other sources may be more or less credible, the agent should adjust its revision of information gained through the social network accordingly (e.g. by integrating perceived trustworthiness and expertise, see Harris et al., 2015). Finally, the agent may wish to update its belief in the reliability of another agent. Other agents may pass on bad information wilfully (lying) or accidentally (incompetence). In either case, the perceived reliability of the source may require adjustment given new evidence. Aside from revising epistemic states, the framework requires semantic categorisation to be able to generate related links between information (semantic networks) and to check if new information contradicts previously held beliefs (epistemic coherence). Finally, the agent must be able to infer new belief states given pre-existing information. This may for example be syllogistic, conditional, or probabilistic. In addition, the agent should be able to draw moral implications from belief-states if the ethical considerations are useful for determining the best way to reach a goal-state.

Information storage As identified in previous unified models, memory is crucial to a standard framework of cognition. An agent must have memory capacities where belief states are stored as epistemic clusters. As humans do not have perfect memory, it is further important to adequately describe memory limitations (degradation, storage capacity, and amount consulted during retrieval) as well as a retrieval function.

Information use Having acquired and managed information, a standard model of cognition needs mechanisms for usage of the information. In order to solve a goal-state, the agent needs to be able to retrieve the encoded and categorised information from the memory system. Further, the agent has to use the information to make a decision. This enables the agent to adapt the information to suit the particular need for achieving the specified problem the agent is attempting to resolve in light of its motivated goal. Additionally, the agent can communicate selected beliefs to other agents. Communicative acts may be truthful or, if the agents deems it sufficiently ethical/beneficial, deceptive. Finally, humans are inherently capable of forward planning. This is a complicated problem, as it relates to potentially multiple goals, the use of different belief-states, and relate to the passage of time in space. Forward planning requires agent predictions, expectations, and goal-states.

Modifying components Humans exhibit differences in the above-mentioned cognitive components such as memory capacity, reasoning strategies, and risk aversion. Consequently, a standard model must be able to express and formally model individual differences. Further, parameters (e.g. emotional states, see Blanchette & Richards, 2010) influence how people solve tasks and reason. A standard model must include parameters that moderate cognitive function or influence the use of information. Finally, the social structure where the human operates is imperative for information input and use. A standard model of cognition must be able to situate the human in different contexts and social scenarios (see the following section), as interaction between agents over time influences how information access, management, and use. Table 1 summarises the broad mechanisms as well as sub-mechanisms of the proposed framework.

Human-environment coupling

The taxonomy of observations and computational models that describes each of these functions points to the need for a standard mode of cognition that is goal-driven and process-oriented. In addition to this problem, Human cognition broadly as well as specifically, though computationally expressible in isolation, occurs in environments where humans engage with the surrounding environment and with other people. While laboratory-based data collection is a justified and valid methodology, it necessarily reduces the complexity of the task by isolating the individual as well as controlling the task stimuli. Here, we suggest that agent-based models are an exciting tool for testing functions and entailments of mechanisms in a way that allows for specified cognitive components in a social system.

While 'sufficiently simple' interactive models may be calculated analytically, some aggregate behavioural pat-

terns are not predictable by analysing each component in isolation and 'multiplying-up' that prediction (Gustafsson & Sternad, 2010). That is, a closed-form solution is not available. Interactions over time and space mean that systems become 'complex' (Johnson, 2007), which are *computationally irreducible*. This means that models of the system cannot rely on closed-form solutions to solve for the state of the system at some point in time without running through all of the intermediate steps. The behaviour of a particular individual cannot be understood without appreciating its interactions and relative position within the system, but the aggregate patterns cannot be produced without the behaviour of each individual. A standard model of cognition requires the capacity for engaging in an adaptive manner with a dynamic environment and with other cognitive agents.

Fundamentally, the need for a generative agent points to Marr's (1982) levels of cognition. He divides cognition in three layers: computational, algorithmic, and physical. While the algorithmic level specifies the necessary procedures required to solve a task and the physical level specifies the hardware limitations, the computational level addresses the broader aim (i.e. *why* the organism approaches the task way). This computational question points to a goal-driven and process-oriented foundation for cognition. We believe the proposed framework (and sub-elements) allow for an agent that is capable of Marr's three levels in a socially oriented, adaptive, and generative manner.

Concluding remarks

We have pointed to the fundamental need for a standard model of cognition. First, if we are to model human cognition *generally*, we need a standard model of cognition that goes beyond modelling of task-specific functions. Second, cognition occurs in complex environments that are frequently computationally irreducible and thus cannot be solved in close-form. This suggests the need for a standard model of cognition that is generative, goal-oriented, and process-driven such that a socially oriented agent can adapt to novel contexts and problems without fundamentally restructuring the cognitive apparatus.

Previous academic work has contributed substantially to the work toward such a standard model, e.g. the ACT-R, SOAR, and Agent_Zero theories. We suggest a broader standard model aimed at socially oriented agents. We have argued that such a model must necessarily include mechanisms for information input, which incorporates goal-oriented motivation, preference ordering, search strategies, and search scope. It must also necessarily include mechanisms for information management concerning belief revision, categorisation, and inference-making. Third, the agent must have a memory system including limitations as well as a retrieval function. Finally, the agent must be able to retrieve and use the information to make more or less long-planned decisions. The framework allows for (heterogeneous) socially oriented agents from which complex social structures can emerge. Further, it allows for adaptive behavioural patterns to emerge individually (e.g. choosing

Table 1: Main components

Information…	Element	Sub-element	Function
1. Input	1.1 Motivated goal	1.1.1 Goal generation	The goal(s) pursued by the agent
1. Input	1.1 Motivated goal	1.1.2 Preference ordering	Importance of sub-states related to the goal(s)
1. Input	1.2 Attention/Perception	1.2.1 Physical perception of the world	Perceptual-motor functions
1. Input	1.2 Attention/Perception	1.2.2 Directed attention	Capacity to direct perceptual-motor functions
1. Input	1.3 Information search	1.3.1 Direct evidence	Evidence required for testing hypotheses about the world
1. Input	1.3 Information search	1.3.2 Indirect evidence	Selection process of other agents to consult
1. Input	1.3 Information search	1.3.3 Search scope	The amount of the world that can be seen at a given time given directed vision
2. Management	2.1 Belief revision	2.1.1 Direct evidence	Integration of new information from direct sources
2. Management	2.1 Belief revision	2.1.2 Indirect evidence	Integration of new information from other sources
2. Management	2.1 Belief revision	2.1.3 Reliability updating	Updating reliability of a source given epistemic coherence and/or quality of advice
2. Management	2.2 Categorisation	2.2.1 Semantic categorisation	Semantic interpretation
2. Management	2.2 Categorisation	2.2.2 Epistemic coherence	Integrating belief-states with previously held belief states to determine if there are internal contradictions
2. Management	2.3 Inferences	2.3.1 Moral judgment	Type of morality (e.g. deontic or utilitarian)
2. Management	2.3 Inferences	2.3.2 Inference strategy	Information use to draw inferences over unknown states
3. Storage	3.1 Memory	3.1.1 Memory system	Type of memory system
3. Storage	3.1 Memory	3.1.2 Memory limitations	Degradation speed, storage space, and amount of memory consulted when making a decision
3. Storage	3.1 Memory	3.1.3 Retrieval function	Specifies how memory is consulted
4. Use	4.1 Decision-making	4.1.1 Decision-making strategy	How to consult information to solve a task to reach a goal
4. Use	4.2 Planning	4.2.1 Forward planning	Capacity to plan ahead (e.g. days or months ahead)
4. Use	4.2 Planning	4.2.2 Discounting (short-term versus long-term gain)	Cost-benefit analysis of temporal discounting
4. Use	4.3 Communication	4.3.1 Communication intention	Whether agents reports the (subjectively held) true or false state of the world to other agents
4. Use	4.3 Communication	4.3.2 Pragmatic enrichment	Pragmatic interpretation

a favoured fishing spot) as well as collectively (e.g. fishing the line). Establishing a standard model with socially oriented, adaptive and generative agents would advance cognitive sciences as well as allow for testing of modular components in novel environments.

Thus, a standard model of cognition would yield significant benefits. First, it would provide a generative and predictive framework that goes beyond computational models of specific issues as well as yield an adaptive and sophisticated agent. Second, implementing the framework in an agent-based model would provide a methodological framework for testing the general applicability of specific mechanism models. Indeed, a standard model of cognition with isolated cognitive functions would necessarily be incomplete without the possibility for interactivity and dynamics between cognitive agents. For instance, it would be possible to test different search strategy models while keeping belief revision and decision-making modules constant. These simulations can then be tested against known or observable data to test the likelihood of the models being apt. Agent-based models would thus be able to test falsifiable predictions from specific components of the general framework. Third, the sub-mechanisms may involve individual differences. This also allows for sensitivity tests of cognitive components that explore which parameters fundamentally alter the overall system. Finally, we believe the overall framework will be beneficial to the development of cognitively realistic AI systems that can be used to simulate market structures, policy interventions, and information flow in social systems.

References

Anderson, J. R. 1983. *The architecture of cognition*. Psychology Press

Blanchette, I: Richards, A. 2010. The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition & Emotion* 24 (4): 561-595

Epstein, J. 2013. *Agent_Zero*: *Toward Neurocognitive foundations For Generative Social Science*. Princeton University Press

Epstein, J. and Axtell, R. 1996. *Growing Artificial Societies*: *Social Science from the Bottom Up*. MIT Press

Gilbert, N. 2008. *Agent-Based Models*. SAGE Publications

Gustafsson, L. and Sternad, M. 2010. Consistent micro, macro, and state-based population modelling. *Mathematical Biosciences* 225: 94-107

Hahn, U., and Oaksford, M. 2007. The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review 114*: 704-732

Harris, A. J. L.: Hahn, U.: Madsen, J. K.: and Hsu, A. S. 2015. The Appeal to Expert Opinion: Quantitative support for a Bayesian Network Approach. *Cognitive Science* 40: 1496-1533

Harris, A.: Hsu, A. and Madsen, J. K. 2012. Because Hitler did it! Quantitative tests of Bayesian argumentation using *Ad Hominem*. *Thinking & Reasoning* 18 (3): 311-343

Johnson, N. 2007. *Simply Complexity: A clear guide to complexity theory*. One World

Kahneman, D. 2011. *Thinking Fast and Slow*. Penguin Books, London: UK

Kemp, C.: Bernstein, A. and Tenenbaum, J. 2005. A generative theory of similarity. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society*.

Koralus, P. and Mascarenhas, S. 2013. The erotetic theory of reasoning: Bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives* 27: 312-365

Laird, J. E. 2012. *The Soar Cognitive Architecture*. MIT Press

Lange, N. D.: Thomas, R. P.: Buttaccio, D. R.: Illingworth, D. A. and Davelaar, E. J. 2013. Working memory dynamics bias the generation of beliefs: The influence of data presentation rate on hypothesis generation. *Psychonomic Bulletin & Review* 20 (1): 171-176

Marr, D. 1982. *Vision*. San Fransisco: Freeman

Miller, J. H. and Page, S. E. 2007. *Complex adaptive systems: An introduction to computational models of social life*. Princeton University Press

Stewart, N.; Chater, N. and Brown, G. D. A. 2006. Decision by sampling. *Cognitive Psychology* 53: 1-26.

Stuhlmüller, A.: Tenenbaum, J. B. and Goodman, N. D. 2010. Learning structured generative concepts. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*

Tenenbaum, J. B.: Kemp, C.: Griffiths, T. L. and Goodman, N. D. 2011. How to Grow a Mind: Statistics, Structure, and Abstraction. *Science* 331 (6022): 1279-1285

Wu, C.: Schultz, E.: Speekenbrink, M.: Nelson, J. and Meder, B. 2017. Mapping the unknown: The spatially correlated multi-armed bandit. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 1357-1362. Austin, TX: Cognitive Science Society,