Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Analogues of mental simulation and imagination in deep learning

Jessica B Hamrick

Mental simulation — the capacity to imagine what will or what could be — is a salient feature of human cognition, playing a key role in a wide range of cognitive abilities. In artificial intelligence, the last few years have seen the development of methods which are analogous to mental models and mental simulation. This paper outlines recent methods in deep learning for constructing such models from data and learning to use them via reinforcement learning, and compares such approaches to human mental simulation. Model-based methods in deep learning can serve as powerful tools for building and scaling cognitive models. However, a number of challenges remain in matching the capacity of human mental simulation for efficiency, compositionality, generalization, and creativity.

Address
DeepMind, 6 Pancras Square, London N1C 4AG, UK

Current Opinion in Behavioral Sciences 2019, **29**:xx-yy

This review comes from a themed issue on **Artificial intelligence**

Edited by **Matt Botvinick** and **Sam Gershman**

https://doi.org/10.1016/j.cobeha.2018.12.011

Introduction

Mental simulation is the ability to construct mental models [1,2] to imagine *what will happen* or *what could be*. Mental simulation is a cornerstone of human cognition [3] and is involved in physical reasoning [4,5], spatial reasoning [6], motor control [7], memory [8], scene construction [9], language [10], counterfactual reasoning [11,12], and more. Indeed, some of the most uniquely human behaviors involve mental simulation, such as designing a skyscraper, performing a scientific thought experiment [13], or writing a novel about people and worlds that do not — and could not — exist. However, such phenomena are challenging to model quantitatively, both because the mental representations used are unclear and because the space of possible behavior is combinatorially explosive.

Artificial intelligence (AI) aims to build agents which are similarly capable of behaving creatively and robustly in novel situations. Perhaps unsurprisingly, there is an analogue to mental simulation in AI: a collection of algorithms referred to as *model-based* methods, with the 'model' referring to a predictive model of what will happen next. While model-based methods have been around for decades [14,15•], recent advances in deep learning (DL) and reinforcement learning (RL) have brought renewed interest in learning and using models. Among the results are systems supporting superhuman performance in games like Go [16,17]. Importantly, these systems must necessarily deal with the computational and representational challenges that have historically faced cognitive modelers.

This paper reviews recent model-based methods in DL and emphasizes where such approaches align with human cognition. The aim is twofold. First, for behavioral scientists, this article provides insight into methods which enable intelligent behaviors in large state and action spaces, with the intent of inspiring future models of mental simulation. Second, for DL and AI researchers, this article compares model-based methods to human capabilities, complementing a number of recent related works [18–21] and clarifying the challenges that lie ahead for building human-level model-based intelligence.

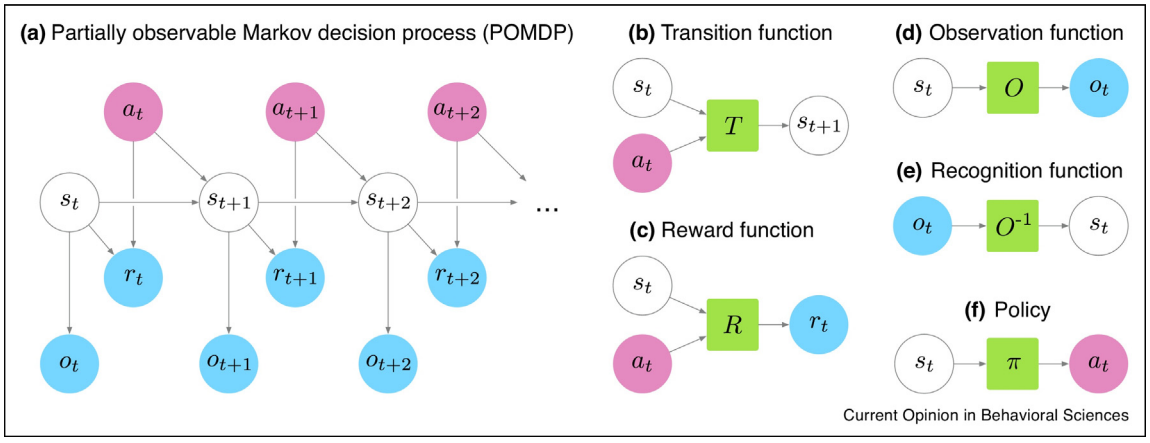
Reinforcement learning

At the core of most model-based methods in DL is the *partially-observable Markov decision process*, or POMDP [22], which governs the relationship between *states* (x), *observations* (o), *actions* (a), and *rewards* (r), as illustrated in Figure 1a. Specifically, these variables are related according to *transition* (T), *observation* (O), *recognition* (O^{-1}), and *reward* (R) functions (Figure 1b–e) as well as a *policy* (π) which produces actions (Figure 1f).

The field of RL is concerned with the problem of finding a policy that achieves maximal reward in a given POMDP. 'Deep' RL implies that the functions in Figure 1 are approximated via neural networks. Much of the research in deep RL is model-free in that it aims to learn policies without knowing anything about the transition, observation, recognition, or reward functions (see [23] for a review). In contrast, model-based deep RL (MBDRL) aims to learn explicit models of these functions which are used to aid in computing a policy, a process referred to as *planning* [15•].

An important component of the POMDP is that of partial observability, in which observations do not contain full

Figure 1



The partially-observable Markov decision process (POMDP). **(a)** A graphical model of the POMDP, where *t* indicates time. A state *s* is a full description of the world, such as the geometries and masses of objects in a scene. An observation *o* is the data that is directly perceived by the agent, such as a visual image. An action *a* is an intervention on the state of the world chosen by the agent, such as 'move left'. A reward *r* is a scalar value that tells the agent how well it is doing on a task and can be likened to the idea of utility or risk. Arrows indicate dependencies between variables. Pink circles indicate variables that can be intervened on; blue indicates variables that are observed; and white indicates variables that are unobserved. **(b–f)** Depictions of the individual functions (in green) that relate variables in the POMDP. The *transition function* (b) takes the current state and action and produces the next state, $s_{t+1} = T(s_t, a)$. The *reward function* (c) takes a state and action and produces a reward (or utility) signal, $r_t = R(s_t, a_t)$. The *observation function* (d) is the process by which sensory data are generated given the current state, $o_t = O(s_t)$. For example, this can sometimes be thought of as a 'rendering' function which produces images given the underlying scene specification. The *recognition function* (e) is the inverse of the observation function, $s_t = O^{-1}(o_t)$, and is analogous to the process of perception. The recognition function is often conditioned on past states and observations (i.e. a 'memory'), to allow aggregation of information across time (for example, velocity, or multiple viewpoints of the same scene). The *policy* (f) is the function which gives rise to actions given the underlying state of the world, $a_t = \pi(s_t)$. The policy is also often conditioned on past memories.

state information. Sometimes, the missing information is minimal (e.g. velocity can be inferred from a few sequential frames); other times, it is severe (e.g. first-person observations are individually not very informative about the layout of a maze). The recognition function thus serves a dual purpose: to infer missing information (e.g. [24,25]), and to transform high-dimensional perceptual observations to a more useful representational format.

The POMDP model provides a useful framing for a variety of mental simulation phenomena, and illustrates how behaviors that seem quite different on the surface share a number of computational similaries (Figure 2). For example, a mental model [1,2] can be seen as a particular latent state representation paired with a corresponding transition function, allowing it to be manipulated or run by via mental actions. The way that people choose *which* mental simulations to run (e.g. the direction of a mental rotation [6]) implies a particular policy and planning method.

Yet, simply framing mental simulation as a POMDP does not tell us where the component functions (Figure 1b–e) come from, what representations they ought to operate over, or how to perform inference in the resulting POMDP. While a number of existing works have answered these questions in simplified observation, state, and action spaces (e.g. [5,26,27]), the answers remain elusive in higher-dimensional settings. Such settings are exactly where DL excels, suggesting that it may prove useful in building cognitive models of mental simulation in richer, more ecologically-valid environments. Indeed, this approach has already been successful in understanding other aspects of the mind and brain, including sensory cortex [28], learning in the prefrontal cortex [29], the psychological representations of natural images [30], and the production of human drawings [31].

Methods in DL for learning models

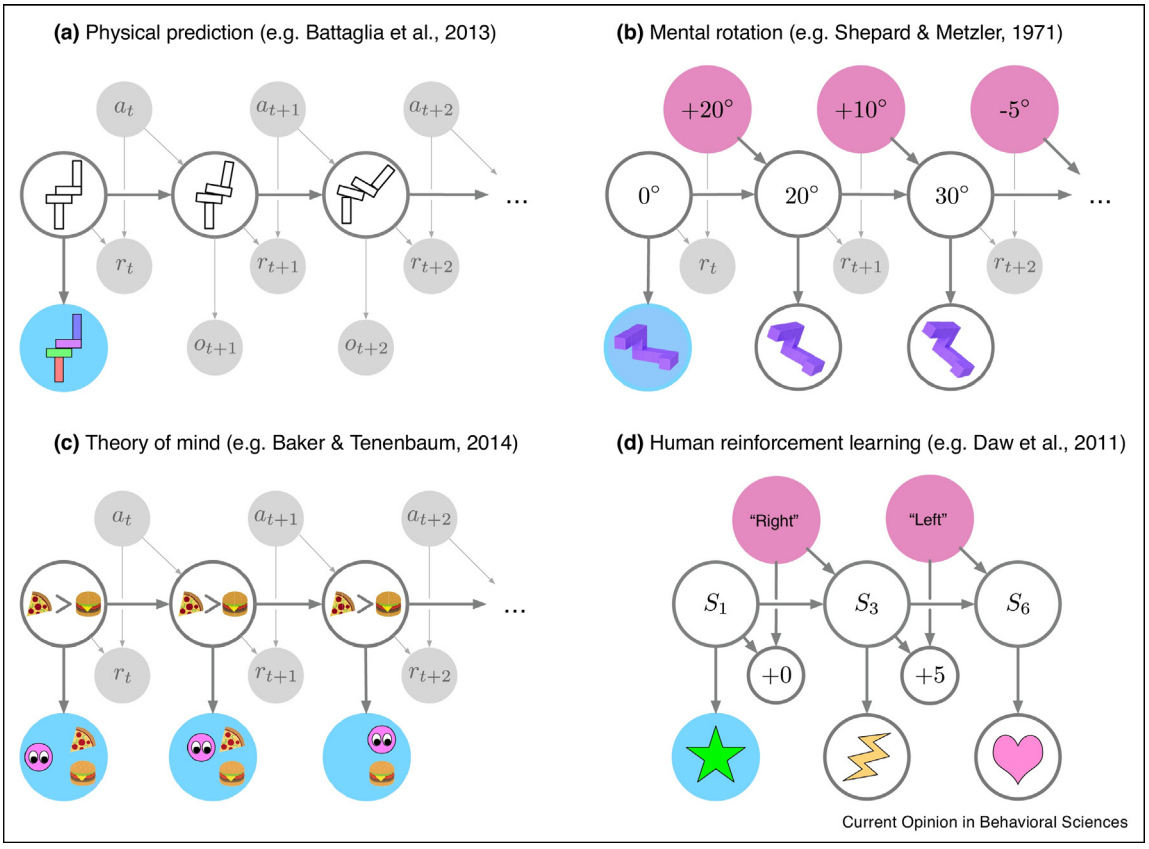
State-transition models

Sometimes, it is assumed that the agent has direct access to a useful representation of the state, obviating the need for a recognition model (Figure 3c). Many approaches to learning such state-transition models use classical recurrent neural networks (e.g. [32,33]). Recently, models which represent the state as a graph [34•] have been used to predict the motion of rigid bodies [35,36], deformable objects [37], articulated robotic systems [38], and multi-agent systems [39].

Observation-transition models

Often, an agent does not have direct access to a useful state representation. One approach to dealing with this issue (often referred to as 'video prediction') learns a transition model directly over sensory observations

Figure 2



Various forms of imagination and mental simulation can be viewed as engaging different aspects of the POMDP model, illustrating how seemingly disparate phenomena are computationally quite similar. In all cases, blue circles indicate relevant observed variables, white circles indicate latent variables, pink circles indicate actions that modify states, and grayed-out circles indicate variables which are not relevant for a particular form of mental simulation. **(a)** Physical prediction tasks like those explored by [5] can be seen as a case where an initial observation is given (e.g. a tower of blocks) and future states are predicted given that observation (e.g. whether the blocks move). **(b)** The mental rotation task from [6], in which it was demonstrated that people imagine objects at different rotations in order to compare them, can be seen as choosing a sequence of actions to produce mental images. **(c)** Theory of mind tasks such as those examined by [26] involve inferring a latent state such as the preferences of another agent (e.g. that the agent prefers pizza over hamburgers) given a sequence of observations (e.g. that the agent picks up a pizza). **(d)** Tasks like the two-step task [27] which probe how humans learn from reinforcement naturally fall under the POMDP paradigm. In such tasks, people must learn to choose actions to navigate through a sequence of symbols in order to maximize a nonstationary reward.

(Figure 3d). For instance, [40] learn a model of pixel motion; [41] predict image masks indicating how a tower of blocks will fall; and [42] predict the boundaries of objects in images.

Prior-constrained latent state-transition models

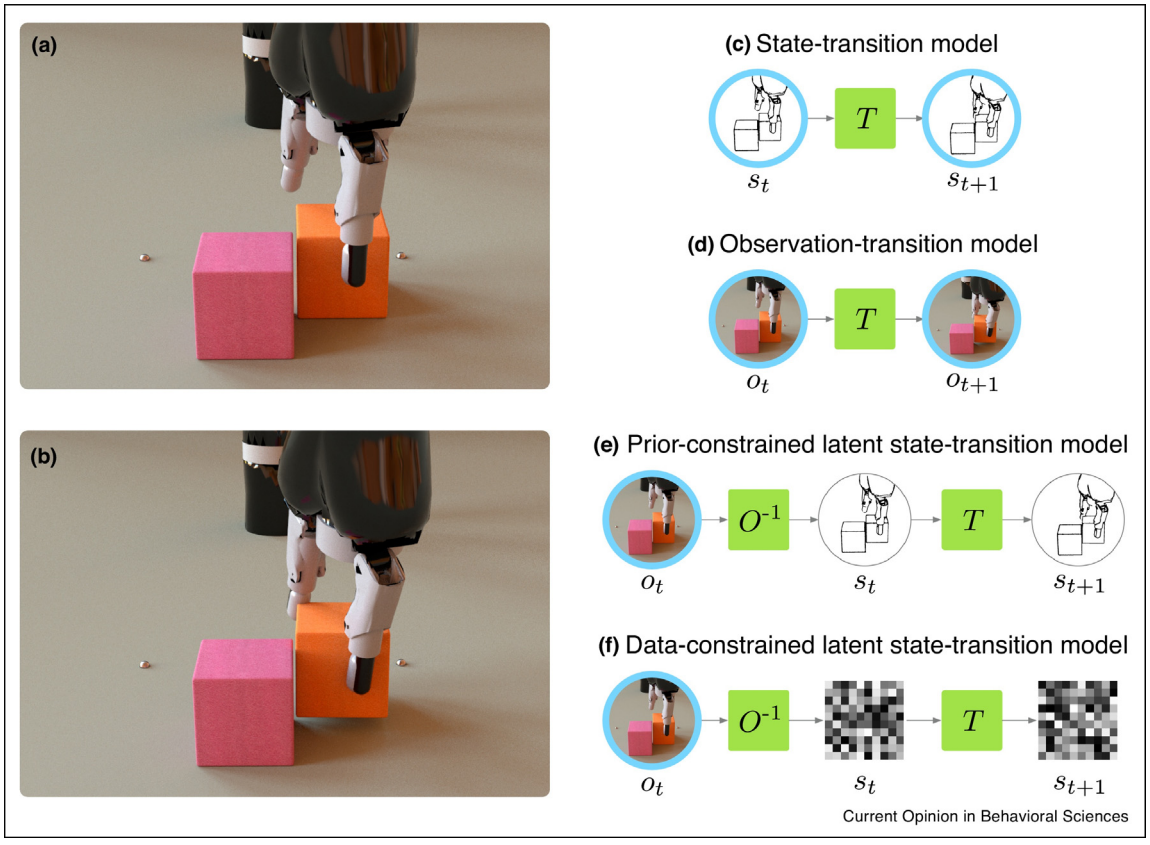
Rather than computing transitions directly over observations, an alternate approach first transforms observations into latent state representations via the recognition function. Sometimes, prior knowledge can be leveraged, resulting in *prior-constrained latent states* (Figure 3e). For example, [43] use pre-existing knowledge about physics to encourage recognition models to recover properties like mass and friction. Other approaches use supervision to train recognition and transition models to predict 2D [44,45] and 3D [46,47•] motion. [48••] use supervision to learn symbolic representations like `on(A,B)`.

Data-constrained latent state-transition models

Another approach infers *data-constrained latent states*, where the representations are influenced more strongly by data than by prior knowledge (Figure 3f). The most common such approach is to find a representation which can be used to predict future observations (e.g. [49–51]). While most approaches assume distributed vector representations, others have explored alternatives such as graphs [52] or low-dimensional binary vectors [53].

Other models have explored different pressures for learning latent state representations beyond reconstructing

Figure 3



Methods for learning models in DL. Most methods focus on learning transition and recognition functions; reward functions are often either assumed to be known or are learned as part of the transition function. Blue outlines indicate variables which are observed. **(a–b)** In a hypothetical scenario, an agent controls a robot arm to pick up a block, and receives pixel-based observations. **(c)** In state-transition models, the underlying states (e.g. the orientation of the blocks and the robot arm) are directly observed. **(d)** In observation-transition models, transitions are learned directly between sensory observations. **(e)** In prior-constrained latent state-transition models, the states must be inferred from observations but often true states are available at training time for supervision, or strong assumptions are made about the dynamics of T or the representation of s. **(f)** In data-constrained latent state-transition models, a latent state is used but no supervision is given over states at training time. The learned latent states are usually distributed and often do not directly correspond to interpretable dimensions such as position, orientation, etc.

observations. For example, one approach is to use policy loss or reward prediction error to shape the latent representations (e.g. [25,54–56]). Because reward is a scalar signal, such representations may not be useful for predicting future observations, but may still be useful for planning. Other objectives include inferring the action taken between observations [57] or maximizing the mutual information between observation-transitions and state-transitions [58•].

Methods in DL for using learned models

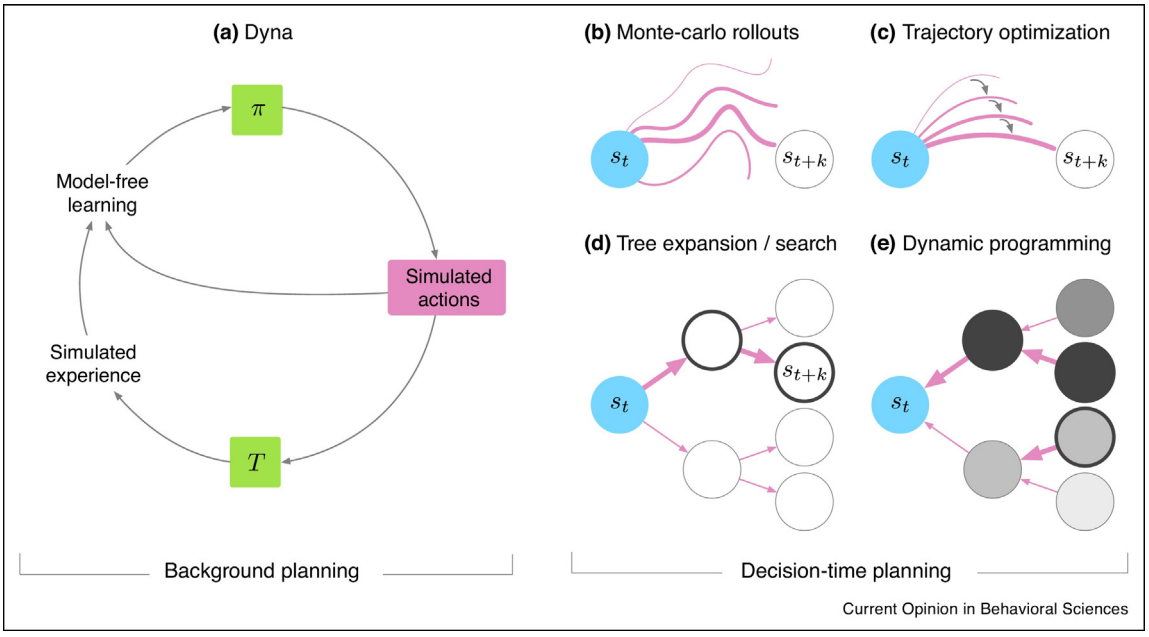
A model on its own does not enable flexible behavior: a planning method is needed to turn predictions into actions. *Background planning* uses models only during the process of learning a model-free policy, while *decision-time planning* uses models during online deliberation [15•].

Background planning

The most popular approach to background planning, Dyna (Figure 4a) [59], uses a model to produce simulated experience in place of real experience [32,53,60]. Other methods backpropagate gradients through learned transition and reward models [61], thus providing more information about an action's utility than a scalar reward does on its own.[1] Another approach uses decision-time planning to improve decisions, and then trains a policy to mimic those decisions [17].

[1] There are close ties between this approach and actor-critic methods, in which a model of future expected reward is learned (the 'critic') and used to optimize the actor via backpropagation. Although actor-critic methods are considered to be model-free because they do not involve a transition model, they are in some sense weakly model-based in that they involve a separate model of the reward which is used to train the policy.

Figure 4



Methods for using models in DL. Pink indicates actions; blue indicates observed states; white indicates unobserved (simulated) states; and green indicates functions. **(a)** Dyna [59] is a background planning method which uses simulated experience in place of real experience. **(b)** Monte-Carlo rollouts, where a base policy is used to sample several sequences of actions and their outcomes (a *trajectory*), after which one is selected according to a criterion such as highest reward. **(c)** Trajectory optimization, where a random sequence of actions is chosen and then optimized to maximize reward. **(d)** Tree expansion or search, in which actions are chosen according to some base policy to search over a tree of possibilities. **(e)** Dynamic programming, in which rewards are computed for the whole state space and then used to recursively compute the maximum future reward for all other states. The shade indicates value, with darker shades indicating high value and lighter shades indicating low value. Readers are referred to [15•] for further details on these methods.

Decision-time planning

One method for decision-time planning simulates *Monte-Carlo rollouts* (Figure 4b) and then chooses the rollout (or trajectory) with highest reward [33]. Alternately, trajectories can be aggregated via a learned mechanism [62]. The choice of base policy for simulating trajectories has varied from random sampling [33]; to approximating the full model-based policy [62]; to learning the base policy end-to-end with the full policy [51].

Another method is trajectory optimization (Figure 4c), which iteratively improves a trajectory. Most works use gradient-based methods [40,54,63], while some have explored using a learned optimization procedure [64,65].

Tree search (Figure 4d) has achieved superhuman performance in Go given a known transition model and learned model-free policy prior [16,17]. Learned models can be used by embedding the tree search into the computation graph itself [56,66•]. Other work learns the decisions that are usually hardcoded into tree search [67••].

Finally, dynamic programming (Figure 4e) performs computations recursively over the entire state space. While techniques like value iteration [14] have classically been used in background planning, recent work incorporates them into DL architectures as decision-time mechanisms [24,68].

Modeling mental simulation with model-based deep RL

The varied approaches to learning and using models in DL have resulted in powerful systems that can model complex physical phenomena [35–38,47•,58•], play difficult puzzle games [62,66•,67••], and control articulated physical systems [33,40,54,63]. But beyond such applications in AI, model-based methods share a number of similarities with human mental simulation (Figure 2), making them an ideal starting point for developing new cognitive models and for scaling existing ones.

Mental imagery

Consider the classic debate regarding which representations underly mental imagery [69–71]. According to the depictive theory (DT) [70], the representations are 2D spatial arrays resembling images. In contrast, the propositional theory (PT) [69] states that the representations are symbolic in nature, without any intrinsic spatial

properties. We can see echoes of these theories in the different structures of transition models. Observation-transition models (Figure 3d) are related to DT in that they operate directly over sensory observations, with intermediate computations operating over 2D convolutional features (e.g. [40]). Prior-constrained latent state-transition models (Figure 3e) may make the assumption that the underlying representation is symbolic (e.g. [48••]), just as PT does. However, neither DT nor PT have strongly considered the role of reward functions or policies [71], in contrast to enactive theories (ET) (e.g. [72]) which consider mental imagery to be strongly coupled to actions. Framing DT, PT, and ET as particular instantiations of the POMDP framework, and modeling them with the tools of MBDRL, provides an avenue for furthering these discussions surrounding mental imagery.

MBDRL may also inform our understanding of mental imagery across the visual [70], auditory [73], and motor modalities [74]. Do these multimodal forms of imagery differ because they deal with different sensory data, or because the underlying mechanisms are themselves also different? MBDRL offers a way to probe this question by training networks with identical or varied architectures on data from different sensory modalities, and comparing the results to human mental imagery phenomena.

Learning by thinking

A longstanding puzzle in cognitive science is that of 'learning by thinking' [75]: how does thought influence behavior without the addition of any new information? One hypothesis proposes that a model-based process trains a model-free action policy [76,77], and has been successfully modeled via Dyna [59] (Figure 4a). However, such work often targets MDPs with small state spaces, which are easier to control experimentally and to compute model predictions for. MBDRL offers the possibility of scaling such theories to behavioral domains with huge state spaces. For example, when combined with DL, such models might also be able to account for the phenomenon of mental practice [78], in which people imagine performing a complex physical action (e.g. throwing a ball) and later exhibit improved performance when actually taking that action.

The control of mental simulation

Finally, an open question is how simulations are controlled during deliberation. An active area of research has investigated the overall choice of whether (and how much) to plan [79,80], treating this choice as a speed-accuracy trade-off and inspiring similarly adaptive approaches in MBDRL [64]. The role of the hippocampus in planning is also an active topic [81,82], with some work suggesting how hippocampal replay might be controlled by a variant of Dyna [77]. Other research has investigated how tree search might support decisions when playing board games [83]. Yet, other domains have received less attention. For example, while people use mental simulation to make predictions about physical scenes like towers of blocks [5], it is unclear how those simulations are engaged when *constructing* towers. Similarly, while mental simulation is used during creative thought (e.g. [84]), it is not well understood which simulations are explored, and why. By casting these problems as POMDPs and solving them with the powerful planning methods from MBDRL, we can produce quantitative, testable hypotheses about how mental simulations might be controlled.

Challenges for model-based deep RL

Model-based deep RL holds the promise of learning rich models of the world from experience and using them to make flexible and robust decisions. However, in comparison to the human capacity for building and running mental models, there are several challenges in fulfilling this promise.

One view of mental simulation holds that it is fast and precise. For example, simulations from forward models in the motor systems must occur in less than 100ms to support real-time action [7]. Similarly, activation of place cells during hippocampal preplay in rats — corresponding to the planning of future trajectories — occurs on the order of 100–300 ms [81]. This view is most consistent with current methods in MBDRL, which require a large number of faster-than-realtime model evaluations before making a decision (e.g. [16,17,33,54]).

Other mental simulations are slow, noisy, and incomplete, with mental simulations lacking full detail [70], exhibiting systematically wrong dynamics [85], and requiring multiple seconds to run [6]. Latent state-transition models have the potential to learn incomplete models of the world, particularly if they do not rely on reconstructing observations. However, almost all planning algorithms assume mostly accurate models. Even in cases where model error is explicitly addressed [32,62–64], it is unclear how well such methods work when the model error is severe. It would seem that the mind can get a lot out of only a handful of incomplete and possibly very inaccurate simulations, a feat which MBDRL methods have yet to achieve.

Mental simulation is also seen as general, flexible, and compositional, supporting behavior across a wide range of different tasks [[38,47•,54]], capturing a large body of commonsense knowledge [[38,47•,54],12], and operating over multiple levels of abstraction [86]. While recent graph network approaches do afford more compositional models than standard RNN approaches [34•], a separate model is still learned for each task or (at best) for a small set of related tasks (e.g. [38,47•,54]). A significant challenge for DL is to build models that seamlessly compose at different levels of abstraction and that are informed by

rich background knowledge about the world, enabling rapid transfer to a diverse range of situations and tasks.

Finally, mental simulation is exploratory, counterfactual, and creative, giving rise to thought experiments [13], children's pretend play [11], and creative works [84]. Mental simulations allow us to conceive of counterfactual worlds that did not come to pass, but which could have [11,12], as well as fully impossible worlds. While the notion of an action-conditional transition model (Figure 1b) does encode some counterfactual knowledge, current methods in DL often struggle to generalize far beyond the scenarios they were trained on [20]. It is an open question of how such methods could entertain concepts as far removed from reality as humans do (such as, 'what if the Earth were replaced by blueberries?' [87]).

While MBDRL holds much promise for building flexible, robust intelligence, it still has a ways to go. To match human cognition, models must be compositional and assembled on-the-fly; methods for planning must succeed with only a handful of evaluations from noisy, incomplete models; and models must be able to generalize far from their training sets, supporting creative exploration and a richer understanding of the world.

Conclusion

The notion of using models of the world to make better decisions has deep roots in the history of both cognitive science [3] and RL [14]. It is unsurprising, then, that both mental simulation and MBDRL share a number of similarities. For cognitive scientists, these similarities suggest that current approaches in MBDRL may be useful starting points for developing new cognitive models and scaling existing models to larger and more complex domains. For DL researchers, they suggest that mental simulation can play an important role in guiding research towards more intelligent agents. In both cases, the integration of model-based methods from DL with theories of mental simulation promises new and exciting research supporting more flexible and creative artificial agents, as well as a deeper understanding of the complexities of the human imagination.

Funding

This work was supported by DeepMind.

Conflict of interest statement

Nothing declared.

Acknowledgements

I would like to thank David Reichert, Peter Battaglia, Aida Nematzadeh, Alex Lerchner, Theo Weber, Kevin Miller, and Kim Stachenfeld for helpful feedback and suggestions, and Tobias Pfaff for both feedback and for rendering the robot images.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- • of special interest

1. Gentner D, Stevens A (Eds): *Mental Models*. Lawrence Erlbaum Associates; 1983.

2. Johnson-Laird PN: **Inference with mental models**. *The Oxford Handbook of Thinking and Reasoning*. 2012:134-145 http://dx.doi.org/10.1093/oxfordhb/9780199734689.001.0001.

3. Craik KJW: *The Nature of Explanation*. 1943.

4. Hegarty M: **Mechanical reasoning by mental simulation**. *Trends Cogn Sci* 2004, **8**:280-285.

5. Battaglia PW, Hamrick JB, Tenenbaum JB: **Simulation as an engine of physical scene understanding**. *Proc Natl Acad Sci U S A* 2013, **110**:18327-18332.

6. Shepard RN, Metzler J: **Mental rotation of three-dimensional objects**. *Science* 1971, **171**:701-703.

7. Wolpert DM, Doya K, Kawato M: **A unifying computational framework for motor control and social interaction**. *Philos Trans R Soc Lond B: Biol Sci* 2003, **358**:593-602.

8. Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK: **The future of memory: remembering, imagining, and the brain**. *Neuron* 2012, **76**:677-694 http://dx.doi.org/10.1016/j.neuron.2012.11.001.

9. Hassabis D, Kumaran D, Maguire EA: **Using imagination to understand the neural basis of episodic memory**. *J Neurosci* 2007, **27**:14365-14374 http://dx.doi.org/10.1523/JNEUROSCI.4549-07.2007.

10. Zwaan RA: **Situation models: the mental leap into imagined worlds**. *Curr Dir Psychol Sci* 1999, **8**:15-18.

11. Harris PL: *The Work of the Imagination*. Blackwell Publishing; 2000.

12. Gerstenberg T, Tenenbaum JB: **Intuitive theories**. In *Oxford Handbook of Causal Reasoning*. Edited by Waldmann M. Oxford University Press; 2017:515-548.

13. Clement JJ: **The role of imagistic simulation in scientific thought experiments**. *Top Cogn Sci* 2009, **1**:686-710 http://dx.doi.org/10.1111/j.1756-8765.2009.01031.x.

14. Bellman R: *Dynamic Programming*. Princeton University Press; 1957.

15. • Sutton RS, Barto AG: *Reinforcement Learning*. edn 2. MIT Press; 2018.
A new version of the canonical text on reinforcement learning, providing a comprehensive guide to the core areas of RL. It also contains an updated, in-depth explanation of the differences between the various planning methods and goes into more detail than what I have presented here.

16. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D: **Mastering the game of Go with deep neural networks and tree search**. *Nature* 2016, **529**:484-489.

17. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D: **Mastering the game of Go without human knowledge**. *Nature* 2017, **550**:354-359.

18. Marblestone AH, Wayne G, Kording KP: **Toward an integration of deep learning and neuroscience**. *Front Comput Neurosci* 2016, **10**:94.

19. Hassabis D, Kumaran D, Summerfield C, Botvinick M: **Neuroscience-inspired artificial intelligence**. *Neuron* 2017, **95**:245-258.

20. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ: **Building machines that learn and think like people**. *Behav Brain Sci* 2017, **40**:1-72 http://dx.doi.org/10.1017/S0140525X16001837.

21. Kunda M: **Visual mental imagery: a view from artificial intelligence**. *Cortex* 2018:155-172.

22. Kaelbling LP, Littman ML, Cassandra AR: **Planning and acting in partially observable stochastic domains**. *Artif Intell* 1998, **101**:99-134.

23. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA: **Deep reinforcement learning: a brief survey**. *IEEE Signal Process Mag* 2017, **34**:26-38.

24. Karkus P, Hsu D, Lee WS: **QMDP-Net: deep learning for planning under partial observability**. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* 2017.

25. Igl M, Zintgraf L, Le TA, Wood F, Whiteson S: **Deep variational reinforcement learning for POMDPs**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.

26. Baker CL, Tenenbaum JB: **Modeling human plan recognition using Bayesian theory of mind**. *Plan, Activity, and Intent Recognition: Theory and Practice*. 2014:177-204.

27. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model-based influences on humans' choices and striatal prediction errors**. *Neuron* 2011, **69**:1204-1215.

28. Yamins DL, DiCarlo JJ: **Using goal-driven deep learning models to understand sensory cortex**. *Nat Neurosci* 2016, **19**:356.

29. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M: **Prefrontal cortex as a meta-reinforcement learning system**. *Nat Neurosci* 2018, **21**:860.

30. Peterson J, Abbott J, Griffiths T: **Evaluating (and improving) the correspondence between deep neural networks and human representations**. *Cogn Sci* 2018:1-35. (in press).

31. Fan JE, Yamins DLK, Turk-Browne NB: **Common object representations for visual production and recognition**. *Cogn Sci* 2018, **42**:2670-2698.

32. Feinberg V, Wan A, Stoica I, Jordan MI, Gonzalez JE, Levine S: **Model-based value expansion for efficient model-free reinforcement learning**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.

33. Nagabandi A, Kahn G, Fearing RS, Levine S: **Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning**. *Proceedings of the International Conference on Robotics and Automation (ICRA 2018)* 2018.

34. • Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R *et al.*: *Relational Inductive Biases, Deep Learning, and Graph Networks*. 2018:1-38arXiv:1806.01261.
A review covering a class of deep learning models called graph networks, which offer a more compositional, object-based representation for learning transition models than do standard LSTM models.

35. Battaglia P, Pascanu R, Lai M, Rezende D, Kavukcuoglu K: **Interaction networks for learning about objects, relations and physics**. *Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS 2016)* 2016.

36. Chang MB, Ullman T, Torralba A, Tenenbaum JB: **A compositional object-based approach to learning physical dynamics**. *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)* 2017.

37. Mrowca D, Zhuang C, Wang E, Haber N, Fei-Fei L, Tenenbaum JB, Yamins DLK: **Flexible neural representation for physics prediction**. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)* 2018.

38. Sanchez-Gonzalez A, Heess N, Springenberg JT, Merel J, Riedmiller M, Hadsell R, Battaglia P: **Graph networks as learnable physics engines for inference and control**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.

39. Hoshen Y: **VAIN: attentional multi-agent predictive modeling**. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* 2017.

40. Finn C, Levine S: **Deep visual foresight for planning robot motion**. *Proceedings of the International Conference on Robotics and Automation (ICRA 2017)* 2017.

41. Lerer A, Gross S, Fergus R: **Learning physical intuition of block towers by example**. *Proceedings of the 33rd International Conference on Machine Learning (ICML 2016)* 2016.

42. Bhattacharyya A, Malinowski M, Schiele B, Fritz M: **Long-term image boundary prediction**. *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18)* 2018.

43. Wu J, Yildirim I, Lim JJ, Freeman WT, Tenenbaum JB: **Galileo: perceiving physical object properties by integrating a physics engine with deep learning**. *Proceedings of the 29th Conference on Neural Information Processing Systems (NeurIPS 2015)* 2015.

44. Fragkiadaki K, Agrawal P, Levine S, Malik J: **Learning visual predictive models of physics for playing billiards**. *Proceedings of the 4th International Conference on Learning Representations (ICLR 2016)* 2016.

45. Watters N, Tacchetti A, Weber T, Pascanu R, Battaglia P, Zoran D: **Visual interaction networks: learning a physics simulator from video**. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* 2017.

46. Mottaghi R, Rastegari M, Gupta A, Farhadi A: **"What happens if . . . ": learning to predict the effect of forces in images**. *Proceedings of the European Conference on Computer Vision (ECCV)* 2016.

47. • Wu J, Lu E, Kohli P, Freeman WT, Tenenbaum JB: **Learning to see physics via visual de-animation**. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* 2017.
A very nice demonstration of how to leverage prior knowledge of physical dynamics and graphical rendering in order to model and control the behavior of complex physical scenes, like towers of blocks.

48. •• Zhang A, Lerer A, Sukhbaatar S, Fergus R, Szlam A: **Composable planning with attributes**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.
A unique approach for learning prior-constrained latent state-transition models, differing substantially from most work in model-based deep RL. This paper learns a transition graph over latent states which are represented propositionally (such as `on(A,B)`), thus abstracting both in space and time.

49. Chiappa S, Racaniere S, Wierstra D, Mohamed S: **Recurrent environment simulators**. *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)* 2017.

50. Ha D, Schmidhuber J: **Recurrent world models facilitate policy evolution**. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)* 2018.

51. Buesing L, Weber T, Racanière S, Eslami SMA, Rezende D, Reichert DP, Viola F, Besse F, Gregor K, Hassabis D, Wierstra D: *Learning and Querying Fast Generative Models for Reinforcement Learning*. 2018:1-15arXiv:1802.03006.

52. van Steenkiste S, Chang M, Greff K, Schmidhuber J: **Relational neural expectation maximization: unsupervised discovery of objects and their interactions**. *Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)* 2018.

53. Corneil D, Gerstner W, Brea J: **Efficient model-based deep reinforcement learning with variational state tabulation**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.

54. Srinivas A, Jabri A, Abbeel P, Levine S, Finn C: **Universal planning networks**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.

55. Silver D, van Hasselt H, Hessel M, Schaul T, Guez A, Harley T, Dulac-Arnold G, Reichert D, Rabinowitz N, Barreto A, Degris T: **The predictron: end-to-end learning and planning**. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)* 2017.

56. Oh J, Singh S, Lee H: **Value prediction network**. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* 2017.

57. Agrawal P, Nair A, Abbeel P, Malik J, Levine S: **Learning to poke by poking: experiential learning of intuitive physics**. *Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS 2016)* 2016.

58. • Kurutach T, Tamar A, Yang G, Russell S, Abbeel P: **Learning plannable representations with Causal InfoGAN**. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)* 2018.

This paper presents an unconventional approach for learning data-constrained state-transition models, by maximizing the mutual information between pairs of sequential observations and pairs of sequential states. This results in a transition function which does not necessarily have to capture full detail of the observations in the latent state, but which must at least capture the general structure of the dynamics.

59. Sutton RS: **Integrated architectures for learning, planning, and reacting based on approximating dynamic programming**. *Proceedings of the 7th International Conference on Machine Learning (ICML 1990)* 1990.

60. Gu S, Lillicrap T, Sutskever I, Levine S: **Continuous deep Q-learning with model-based acceleration**. *Proceedings of the 33rd International Conference on Machine Learning (ICML 2016)* 2016.

61. Heess N, Wayne G, Silver D, Lillicrap T, Tassa Y, Erez T: **Learning continuous control policies by stochastic value gradients**. *Proceedings of the 29th Conference on Neural Information Processing Systems (NeurIPS 2015)* 2015.

62. Weber T, Racanière S, Reichert DP, Buesing L, Guez A, Rezende D, Badia AP, Vinyals O, Heess N, Li Y, Pascanu R, Battaglia P, Silver DHD, Wierstra D: **Imagination-augmented agents for deep reinforcement learning**. *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)* 2017.

63. Chua K, Calandra R, McAllister R, Levine S: **Deep reinforcement learning in a handful of trials using probabilistic dynamics models**. *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)* 2018.

64. Hamrick JB, Ballard AJ, Pascanu R, Vinyals O, Heess N, Battaglia PW: **Metacontrol for adaptive imagination-based optimization**. *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017)* 2017.

65. Pascanu R, Li Y, Vinyals O, Heess N, Buesing L, Racanière S, Reichert D, Weber T, Wierstra D, Battaglia P: *Learning Model-based Planning from Scratch*. 2017:1-13arXiv:1707.06170.

66. • Farquhar G, Rocktäschel T, Igl M, Whiteson S: **TreeQN and ATreeC: differentiable tree planning for deep reinforcement learning**. *Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)* 2018.

An elegant way to incorporate tree search into a deep RL architecture using learned abstract state-transition models, demonstrating even a small amount of search can improve upon standard DQN and A2C baselines.

67. •• Guez A, Weber T, Antonoglou I, Simonyan K, Vinyals O, Wierstra D, Munos R, Silver D: **Learning to search with MCTSnets**. *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* 2018.

This paper proposes a method for learning how to perform a tree search (rather than using handcrafted decisions about which states to explore during search, as is normally done) and demonstrates significantly improved search efficiency in Sokoban. Although this method still requires more model evaluations than humans seem to need, it is a promising step in the direction of more efficiently using simulation.

68. Tamar A, Wu Y, Thomas G, Levine S, Abbeel P: **Value iteration networks**. *Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS 2016)* 2016.

69. Pylyshyn ZW: **Mental imagery: in search of a theory**. *Behav Brain Sci* 2002, **25**:157-182 http://dx.doi.org/10.1017/S0140525X02000043.

70. Kosslyn SM, Thompson WL, Ganis G: *The Case for Mental Imagery*. Oxford University Press; 2006.

71. Thomas NJ: **Mental imagery**. In *The Stanford Encyclopedia of Philosophy,* Spring 2018 edition. Edited by Zalta EN. Metaphysics Research Lab, Stanford University; 2018.

72. Grush R: **The emulation theory of representation: motor control, imagery, and perception**. *Behav Brain Sci* 2004, **27**:377-396.

73. Zatorre RJ, Halpern AR: **Mental concerts: musical imagery and auditory cortex**. *Neuron* 2005, **47**:9-12.

74. Jeannerod M: **Mental imagery in the motor context**. *Neuropsychologia* 1995, **33**:1419-1432.

75. Lombrozo T: *"Learning by Thinking" in Science and in Everyday Life*. Oxford University Press; 2017:1-26.

76. Gershman SJ, Zhou J, Kommers C: **Imaginative reinforcement learning: computational principles and neural mechanisms**. *J Cogn Neurosci* 2017, **29**:2103-2113.

77. Mattar MG, Daw ND: **Prioritized memory access explains planning and hippocampal replay**. *bioRxiv* 2018. 225664.

78. Driskell JE, Copper C, Moran A: **Does mental practice enhance performance?** *J Appl Psychol* 1994, **79**:481.

79. Keramati M, Dezfouli A, Piray P: **Speed/accuracy trade-off between the habitual and the goal-directed processes**. *PLoS Comput Biol* 2011, **7**:e1002055.

80. Keramati M, Smittenaar P, Dolan RJ, Dayan P: **Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum**. *Proc Natl Acad Sci U S A* 2016, **113**:12868-12873.

81. Ólafsdóttir HF, Barry C, Saleem AB, Hassabis D, Spiers HJ: **Hippocampal place cells construct reward related sequences through unexplored space**. *eLife* 2015, **4**:e06063 http://dx.doi.org/10.7554/eLife.06063.001.

82. Miller KJ, Botvinick MM, Brody CD: **Dorsal hippocampus contributes to model-based planning**. *Nat Neurosci* 2017, **20**:1269.

83. van Opheusden B, Bnaya Z, Galbiati G, Ma WJ: **Do people think like computers?** *Proceedings of the 9th Annual Conference on Computers and Games* 2016 http://dx.doi.org/10.1007/978-3-319-50935-820.

84. Finke RA, Slayton K: **Explorations of creative visual synthesis in mental imagery**. *Mem Cogn* 1988, **16**:252-257.

85. McCloskey M: **Intuitive physics**. *Sci Am* 1983, **248**:122-131.

86. Dezfouli A, Balleine BW: **Actions, action sequences and habits: evidence that goal-directed and habitual action control are hierarchically organized**. *PLoS Comput Biol* 2013, **9**:e1003364.

87. Sandberg A: *Blueberry Earth*. 2018:1-7arXiv:1807.10553.