Meaning Versus Information, Prediction Versus Memory, and Question Versus Answer

Yoonsuck Choe

Department of Computer Science and Engineering, Texas A&M University

Abstract

Brain science and artificial intelligence have made great progress toward the understanding and engineering of the human mind. The progress has accelerated significantly since the turn of the century thanks to new methods for probing the brain (both structure and function), and rapid development in deep learning research. However, despite these new developments, there are still many open questions, such as how to understand the brain at the system level, and various robustness issues and limitations of deep learning. In this chapter, I will talk about some of the concepts that are central to brain science and artificial intelligence, such as information and memory, and discuss how a different view on these concepts can help us move forward, beyond current limits of our understanding in these fields.

1 Introduction

Brain and neuroscience, psychology, artificial intelligence, all strive to understand and replicate the functioning of the human mind. Advanced methods for imaging, monitoring, and altering the activity of the brain at the whole-brain scale are now available, allowing us to probe the brain in unprecedented detail. These methods include high-resolution 3D imaging (using both physical [serial] sectioning and optical sectioning), monitoring ongoing neural activity (e.g. calcium imaging), and selective activation of genetically specific neurons (optogenetics). On the other hand, in artificial intelligence, deep learning based on decades-old neural networks research made exponential progress, and it is now routinely beating human performance in many areas including object recognition and game playing.

However, despite such progress in both fields, there are still many open questions. In brain science, one of the main questions is how to put together the many detailed experimental results into a system-level understanding of brain function. Also, there is the ultimate question to understand the phenomenon of consciousness. In artificial intelligence research, especially in deep learning, there are lingering issues of robustness (for example, deep neural networks are easily fooled by slightly altered

inputs such as adversarial inputs), and interpretability (deep neural networks are basically a black-box, and humans do not understand why or how they work so well).

In this essay, I will talk about some of the concepts that are central to brain science and artificial intelligence, such as information and memory, and discuss how a slightly different view on these concepts can help us move forward, beyond current limits of our understanding in these fields.

The rest of this chapter is organized as follows. In section 2, I will discuss meaning vs. information, with an emphasis on the need to consider the sensorimotor nature of brain function. In section 3, I will talk about prediction and memory, in the context of synaptic plasticity and brain dynamics. In section 4, I will move on to a broader topic of question vs. answer, and discuss how this dichotomy is relevant to both brain science and artificial intelligence. Section 5 will include some further discussions, followed by conclusions (section 6).

2. Meaning vs. Information

The concepts of computing and information have fundamentally altered the way we think about everything, including brain function and artificial intelligence. In analyzing the brain and in building and interpreting intelligent artifacts, computing has become a powerful metaphor, and information has become the fundamental unit of processing and measurement. We think about the brain and artificial neural networks in terms of computing and information processing, and measure their information content. In this section, I will talk mostly about information.

First of all, what is information? We tend to use this word in a very loose sense, and information defined in this way is imbued with meaning (or semantic content). In a scientific/engineering discussion, information usually refers to the definition given in Claude Shannon's information theory [1]. In information theory, information is based on the probability of occurrence of each piece of message, and the concept is used to derive optimal bounds on the transmission of data.

However, there can be an issue if we try to use Shannon's definition of information in our everyday sense, since as Shannon explicitly stated in his primary work on information theory, information defined as such does not have any meaning attached to it. So, for example, in an engineered information system, when we store information or transmit information, the data themselves do not have meaning. The meaning only resides in the human who accesses and views the information. All the processing and transmission in the information system is at a symbolic level, not at a semantic level.

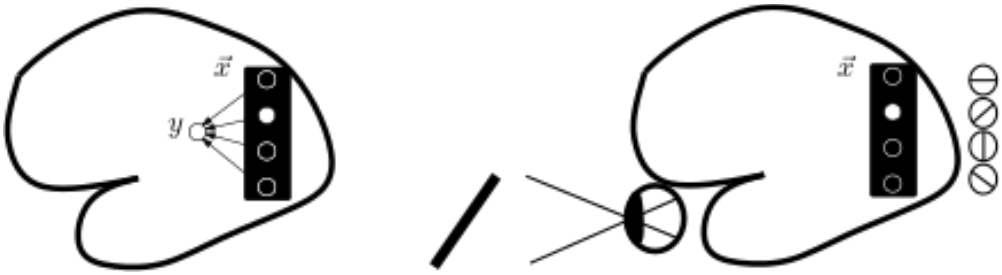
Philosopher John Searle's paper on the "Chinese room argument" [2] made clear of the limitation of the computational/information processing view of cognition. Inside the Chinese room there is an English monolingual person with all the instructions for processing Chinese language. The room has two mail slots, one for input, and one for output. A Chinese speaker standing outside the room writes down something (in Chinese) on a piece of paper and deposits it in the input slot, and the English speaker inside will process the information based on the instructions present in the room and meticulously draw (not write) the answer on a piece of paper and returns it through the output slot. From the outside, the Chinese room speaks and understands perfect Chinese, but there is no true understanding of Chinese in this system. The main problem is that the information within the room lacks meaning, and this is why grounding is necessary; grounding in the sense that information is grounded in reality, not hovering above in an abstract realm of symbols (see Stevan Harnad's concept of symbol grounding [3]). Many current artificial intelligence systems including deep learning tend to lack such grounding, and this can lead to brittleness, since these systems simply learn the input output mapping, without understanding.

Thus, we need to think in terms of the meaning of the information, how semantic grounding is to be done: How does the brain ground information within itself? (see [4] for a similar approach, the "inside-out" approach, in neuroscience) How can artificial intelligence systems ground information within itself? What is the nature of such grounding? Perceptual? Referential? This can be a very complex problem, so let us consider a greatly simplified scenario.

Figure 1 Inside a black box. There are four lights that blink at different times. What do they mean?

Suppose you are sitting inside a totally dark room, and you only observe the occasional blinking of some light bulbs (Figure 1). You count the bulbs, and it looks like there are four of them. Each of these bulbs seem to be representing some information, but you are unsure what they mean. So, here you have a classic symbol grounding problem. The light blubs are like symbols, and they represent something. However, sitting inside this room, it seems that there is no way you can figure out the meaning of these blinking lights. Now consider that the dark room is the primary visual cortex (V1), and the four

light bulbs are neurons that represent something, and in place of you, we put inside the room the downstream visual areas (Figure 2a). With our reasoning above, it would suggest that the downstream visual areas have no way to understand the meaning of V1 activities. Paradoxically, the problem seems to be trivial, if seen from the outside (Figure 2b). This is absurd, since humans have no problem with visual understanding, even without any knowledge of how the orientation-tuned neurons in the visual cortex



(a) Brain, from the inside (b) Brain, from the outside

works!

Figure 2. Brain, from the inside and from the outside. (a) From the inside, the neural activity in the simple visual cortex $\vec{x}$ cannot be deciphered by the downstream neuron *y*, just by looking at $\vec{x}$. (b) From the outside, by presenting different input stimuli in the environment (e.g., a 45º diagonal line), and observing which neurons are activated in $\vec{x}$ (e.g. the second neuron from the top) when that input is presented, we can tell what the neurons in $\vec{x}$ represent (the four angles shown to the right).

It turns out that this problem can only be solved if we allow motor interaction from within the system. Inside the dark room, we can install a joystick connected to an external camera, and the person sitting inside can move it around and see how the joystick movement relates to the changes in the blinking light in a systematic manner. Consider the case where the four light bulbs represent four different orientations 0º, 45º, 90º, and 135º, respectively. How can movement of the joystick reveal the meaning of these light bulbs? We can map this scenario into the brain (Figure 2a), and consider the joystick as the eye movement signal in the motor cortex (Figure 3).

Let us first consider what is happening inside the brain (Figure 3a, “Inside”, top row). The second neuron is turned on, and the eye movement (dashed arrow) is along 45 º. In the next step (Figure 3a, “Inside”, bottom row), the eye movement is along 225 º. The two states can alternate, and curiously the second neuron keeps on firing and all other neurons are silent. Now, let us see what’s happening in the environment. There is a long 45 º line in the external environment, and the gaze (circle) is tracing along this line in the 45 º direction (Figure 3a, “Outside”, top row). In the next step (Figure 3a, “Outside”, bottom row), the gaze is tracing backward in the 225 º direction. Of course, this “outside” perspective is not known to the observer “inside”. However, by discovering a specific type of action (45 º to 225 º back-and-forth movement) that keeps the internal state invariant (the second neuron), the property of the internal representation (45 º orientation) can be inferred.

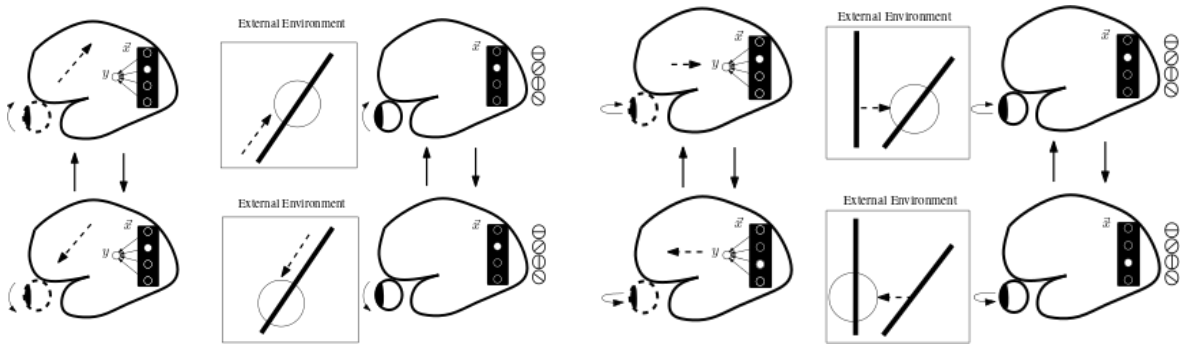


Figure 3. Changes in the internal representation due to action (eye gaze). (a) Unchanging internal state when action and stimulus has the same property. (b) Changing internal state when action and stimulus are misaligned.

This kind of internal state invariance breaks down when the property of the action is not aligned with the property of the internal representation, as we can see in Figure 3b. In this case, back-and-forth horizontal movement (top row to bottom row, etc.) leads to alternating internal states (second neuron is active, then the third neuron is active, and the same sequence repeats).

Through this kind of sensorimotor exploration, the property of the internal representation can be recovered, from within the system without direct perceptual access to the external environment, thus the meaning can remain intrinsic to the system.

In our lab, we explored these ideas in a reinforcement learning setting (learn a policy *p* that maps from state *S* [orientation] to action *A* [gaze direction]), where we showed that the internal state invariance criterion can be used as a reward for motor grounding of internal sensory representations in a simple visuomotor agent. See [5] and subsequent works for more details.

To sum up, meaning is central to brain science and artificial intelligence, and to provide meaning to information (i.e., grounding), it is critical to consider the sensorimotor aspect of the information system, whether natural or artificial.

3. Prediction vs. Memory

Many questions in brain and neuroscience focus on the concept of plasticity, how the brain changes and adapts due to experience, and this leads to the question of memory.

Connections between neurons adapt over time (synaptic plasticity: long term, short term, etc.), and ongoing neural dynamic of the brain can also be altered by the immediate input stimulus. On a higher level, plasticity is usually considered in relation to various forms of memory: long term memory, short term memory, working memory, episodic memory, implicit memory, explicit memory, etc. Also, in a commonsense way, people ask how the brain remembers, and what constitutes memory in the brain. In artificial intelligence, the same is true: How information should be represented, stored, and retrieved; how connection weights should be adapted in artificial neural networks to store knowledge; and how neural networks can be used to utilize external memory, etc.

What is memory, and how is it related to prediction, and why should we think more about prediction than memory? Memory is backward looking, directed toward the past, while prediction is forward looking, and is directed toward the future. Memory enables prediction, since without memory, the system will be purely reactive, living in the eternal present. So, again, why should we direct our attention toward prediction? In terms of brain function and artifacts that try to mimic it, prediction is of prime importance. In our everyday life, moment to moment prediction and long term prediction play a critical role. Simple tasks as walking, navigating, and many daily activities involving interaction with the environment and with other humans require prediction. Long term predictions of phenomena such as seasonal changes enable planning and improved productivity. So, in a sense, prediction is an important brain function, and it is increasingly being recognized as a central function of the brain as well as a key ingredient in intelligent machines (for an overview of related ideas, see Andy Clark's paper [6], and various papers on the use of predicted future states in reinforcement learning [7]).

In this section, I will talk about how such predictive function could have emerged in the brain, how it is related to synaptic plasticity mechanisms (memory), how it is relevant to the study of neural networks, and how predictive properties in the brain can be linked to higher level phenomena such as consciousness.

First, consider delay in the nervous system. Neurons send their signals to their receiving neurons via elongated wires called axons. Transmission through these axons can take few milliseconds (ms), the duration depending on various factors such as the length, diameter, and whether the axon is insulated with myelin or not. When you add up the delay, it comes to a significant amount of time: about 180 ms to 260 ms from stimulus presentation to behavioral reaction [8]. This kind of delay may be considered bad for the system, since it can be a matter of life and death, especially for fast moving animals. Also, in engineering systems, delay is considered a great hindrance. However, delay can be useful in two ways: (1) in a reactive system such as a feedforward neural network, addition of delay in the input can effectively add memory, and (2) mechanisms

evolved to counteract the adverse effects of delay can naturally lead to predictive capabilities.

In [9], we showed that addition of delay in feedforward neural network controller can solve a 2D pole balancing problem that does not include velocity input. Also, in a series of works we showed that certain forms of synaptic plasticity (dynamic synapses) can be considered as a delay compensation mechanism, and how it relates to curious perceptual phenomena such as the flash lag effect (see [10] for an overview). In flash lag effect, a moving object is perceived as being ahead of a statically flashed object that is spatially aligned (Figure 4). One explanation for this phenomenon is that the brain is compensating for the delay in its system, by generating an illusion that is aligned, in real time, with the current state of the external environment. For example, image of the two aligned objects hit the retina (Figure 4, *t=3*). The information takes several milliseconds to reach the visual area in the brain (Figure 4, *t=4*, bottom row). In the meanwhile, the gray object moves ahead (Figure 4, *t=4*, top row, gray box), so by the time the two objects are perceived (when the information arrives in the visual area at *t=4*), in the environment, they are misaligned because the moving object has moved on. The argument is that flash-lag effect allows the brain to perceive this as misaligned objects, which is more in line with the actual environmental state at the time of perception (*t=4*). We showed that facilitating neural dynamics, based on dynamic synapses (the facilitating kind, not the depressing kind: see Markram and colleagues' works cited in [10]) can replicate this phenomenon, and furthermore, the use of spike-timing-dependent plasticity (STDP) can help explain more complex phenomena such as orientation flash-lag effect (see [10] and references within).

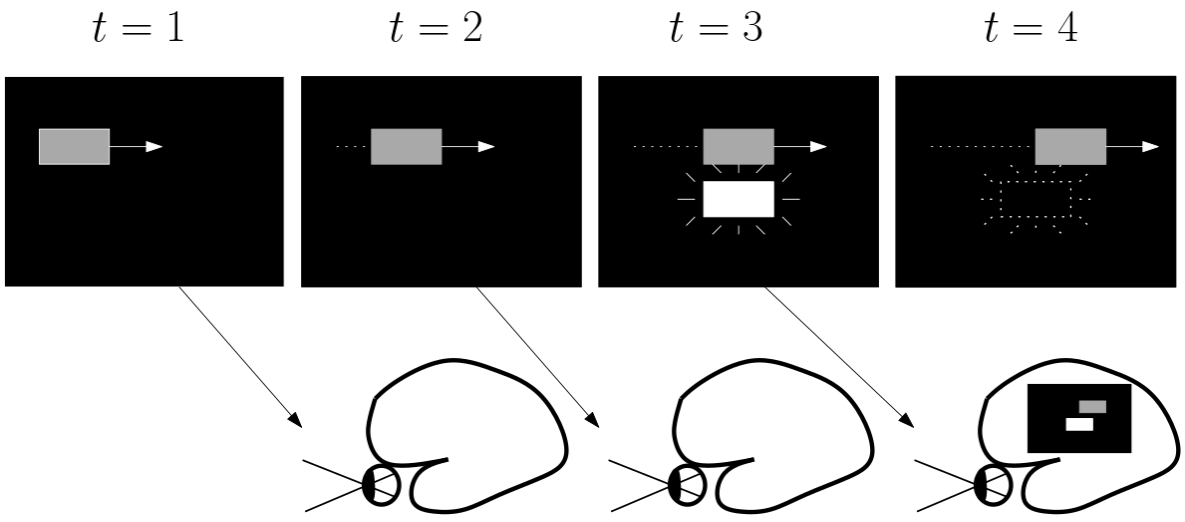


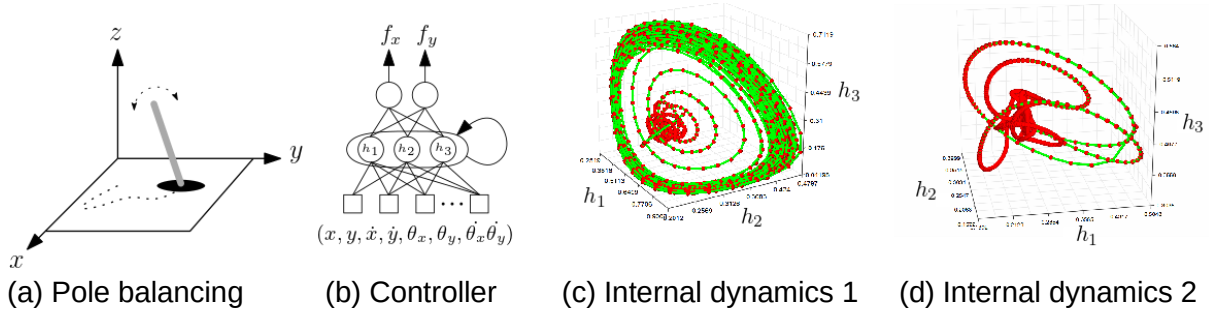
Figure 4. Flash Lag Effect. The white bar is flashed briefly (*t=3*) when the moving bar (gray) passes its location. This event is perceived as staggered (as shown inside the brain in the bottom right, *t=4*), not vertically aligned (as shown in *t=3*).

Second, we will consider predictive properties in brain dynamics and how it can be related to higher level phenomena such as consciousness. As we discussed above, prediction seems to be a key function of the brain. How can it also be used to gain insights into phenomena such as consciousness? In consciousness studies, the neural correlate is highly sought after, where neural correlates of consciousness refer to the "... neural events and structures … sufficient for conscious percept or conscious memory" [11]. This view is somewhat static (of course it depends on the definition of "event"), and its dependence on sufficient conditions can lead to issues relating to the hard problem of consciousness — how and why it "feels" like it. In our view, it would be better to first consider the necessary conditions of consciousness, and this led us to the realization that the property of brain dynamics, not just isolated "events", need to be considered. We also found that predictive property in brain dynamics has an important role to play in consciousness, and this is how the discussion of consciousness comes into the picture in this section [12].

Let us consider necessary conditions of consciousness. We begin by considering consciousness and its subject. There cannot be a consciousness without a subject, since consciousness, being a subjective phenomenon, cannot be subjective without a subject (see [13] pp 191-193, on how sensations "belonging to a subject" is a major property of consciousness). Next, consider the property of the subject (or let us say "self"). Self is the author of its own actions, and there is a very peculiar property about these actions authored by the self — that it is 100% predictable. When I say "I will clap my hands in 5 seconds", in 5 seconds, I will make sure that happens, so that my behavior in such a case is 100% predictable, by myself but not by others. This is quite unlike most phenomena in the world that is not so much the case. In order to support such a prediction, some part of the brain has to have a dynamic pattern that has a predictable property. That is, based on past activation patterns in the neural dynamic trajectory, it needs to be possible to predict the current activation pattern. This, we believe, is an important necessary condition of consciousness (see [12] for details).Through computational simulations and secondary analysis of public EEG data, we showed that predictive dynamics can emerge and have fitness advantage in synthetic evolution [12], and conscious states such as awake condition and REM sleep condition exhibit more predictive dynamics than unconscious states (slow-wave sleep) [14].

For the first study [12] [15], we evolved simple recurrent neural network controllers to tackle the 2D pole-balancing task (Figure 5), and found that successful individuals have a varying degree of predictability in its internal dynamics (how the hidden unit activities change over time, and how predictable the future state is, given a short time window in the past states: Figure 5c-d). This is discouraging, since if individuals with internal

dynamics with high (Figure 5c) or low predictability (Figure 5d) are equally good in behavioral performance, predictive dynamics may not evolve to dominate. However, a slight change in the environment made individuals with high predictive dynamics to proliferate. The only change necessary was to make the task a little harder (make the initial random tilt of the pole to be more). This suggests that predictive internal dynamics have a fitness advantage when the environment changes over time, and this happens to be how the nature is, thus predictive dynamics will become dominant. Not the strongest or the fastest species survive: The most adaptable species survive, where prediction seems to be the key, and this helps satisfy the necessary condition for consciousness.



(a) Pole balancing (b) Controller (c) Internal dynamics 1 (d) Internal dynamics 2

Figure 5. 2D Pole Balancing. (a) The 2D pole balancing task. (b) Recurrent neural network controller. Inputs are the location of the cart and angles of the pole and their respective velocities, and the outputs are the force to be applied in the *x* and the *y* direction (c) Internal dynamics example 1 (hidden states h_1, h_2, h_3 plotted over time in 3D). (d) Internal dynamics example 2. Both (c) and (d) are from successful controllers. (c) and (d) adapted from *[15]*, by permission from IEEE.

In the second study [14], we analyzed publicly available brain EEG data collected during awake, rapid eye movement (REM) sleep, and slow-wave sleep. Since awake and vivid dreaming (REM sleep) are associated with consciousness while deep sleep (slow-wave sleep) with unconsciousness, we measured the predictability in these EEG signal wave forms. We preprocessed the raw EEG signal, computed the inter-peak interval (IPI), the time distance between peaks in the EEG signal, and measured how easy it is to predict the next IPI based on previous IPI data points. We found that awake and REM EEG signals have higher IPI predictability than that of slow-wave sleep, suggesting that IPI predictability and consciousness may be correlated.

In this section I discussed how synaptic plasticity mechanisms can be directly linked to prediction, how delay in the nervous system may have led to predictive capabilities, and how predictive dynamics can serve as a precursor of consciousness. In sum, prediction is a key aspect of the brain, and it should also be considered when designing intelligent artifacts.

4. Question vs. Answer

In both brain science and artificial intelligence, the general focus is to understand how the brain solves problems relating to perceptual, cognitive, and motor tasks, or how to make artificial intelligence algorithms solve problems in vision, natural language processing, game playing, robot control, etc. That is, we are focused on mechanisms that produce answers, and less on mechanisms that pose the questions. Of course we know the importance of asking the right questions, and any researcher is well aware of the importance of picking the right research question. Often times, research involves finding new ways to conceive of the problem, rather than finding new ways of solving problems as conceived [16], and this is especially essential when the conceived problem itself is ill-formed so as to be unsolvable (e.g., “how can we prove Euclid’s 5th postulate” [unsolvable], vs. “can we prove Euclid’s 5th postulate” [solvable]).

In 2012, Mann and I discussed in [17] the need to start paying attention to problem posing, as opposed to (or in addition to) problem solving, in artificial intelligence. It turns out that problem posing has been an active topic in the education literature (see [18] and many subsequent publications), and these works show that learning and problem posing are intricately related. However, this angle is not explored much in artificial intelligence, except for rare exceptions, and I strongly believe integrating learning and problem posing can lead to a much more robust and more general artificial intelligence. Some of those rare exceptions is Schmidhuber’s study, which explicitly addresses this issue. In his Powerplay algorithm, both problems and solvers are parameterized and the algorithm seeks specific problems that are solvable with the current capability of the solver, and loop through this to train an increasingly general problem solver [19]. More recently, question asking has been employed in interactive problem solving in robotics [20] and vision problems [21]. However, these are done within a strict task framework, so open-ended questions or questions that question the validity of existing questions cannot be generated. See [21] for a bit more open-ended approach called inverse Visual Question Answering. More recent works include goal generation and goal selection, where agents come up with their own goals and select from the candidate goals [22] [23]. Some ideas we discussed in [17] for problem posing include: (1) recognizing events in the environment that can be potentially become a problem to be solved, (2) checking if existing problems are ill-posed, and (3) given an overarching goal, come up with smaller problems that may be of different kind than the original goal (if they are of the same kind, straight-forward divide-and-conquer algorithms can be used). For an intelligent agent posing new questions, inventing new tasks, and creating new goals will become increasingly important, as the current learning algorithms cannot easily go beyond its defined task context.

How can the idea of question vs. answer be relevant to brain science? I think it is relevant since the topic has not received attention that it deserves, and question asking (or problem posing) is an important function of the brain. There are many papers on decision making, but not much on how the brain asks questions that requires subsequent decision making. Understanding the brain mechanism of question asking can lead to new discoveries regarding core brain function, and in turn the insight can help us build better intelligent artifacts.

In sum, question asking needs more attention from brain science and artificial intelligence, in order for us to gain a deeper understanding of brain function and to build more intelligent artifacts. Furthermore, we can ask a meta-question: Are we asking the right questions when trying to solve the problem of intelligence, artificial or natural? The pairs of contrasting concepts I discussed in this chapter are largely based on asking this meta question.

5. Discussion

In this chapter, I talked about several dichotomies of concepts that are important to brain science and artificial intelligence: meaning vs. information, prediction vs. memory, and question vs. answer, with an emphasis on the first concept in each pair. Below, I will discuss additional related works that have relevance to the topics I discussed in the preceding sections.

In terms of meaning, deep neural network research commonly approaches the issue from a different angle than the one presented in this chapter — that of embedding, e.g., word and sentence embedding [24] (see [25] for an extensive review on word meaning in minds and machines, which includes a detailed survey on word embedding). The main idea of embedding is to map words or sentences (or other raw input) into a vector space where concepts can be manipulated algebraically based on their meaning. For example, when vector representations of “Germany” and “capital” are added, the resulting vector represents “Berlin” (example from [24]). The main idea in this case is to train a neural network to map the input word to a series of words that appear before or after the input word in a sentence. The hidden layer representation then tends to take on this desired semantic property. This is one step toward meaningful information. However, whether the meaning in this case is intrinsic to the system, i.e., interpretable from within the system, is an open question. A large body of work on huge language models that appeared in recent years based on transformers such as GPT-3 [26], although very powerful, may still suffer from lack of meaning.

As we saw in section 2, the motor aspect is important for semantic grounding of meaning. In a related context, philosopher Ludwig Wittgenstein proposed that the meaning of language is in its use [27]. Basically, this is a departure from meaning based on what it (e.g., a word) represents (see [28] for various views on representation in neuroscience). A more recent thesis in this general direction comes from Glenberg and Robertson [29], where they emphasized that "what gives meaning to a situation is grounded in actions particularized for that situation", thus taking an action-oriented view of grounding (on a more general note, see [30] [31] [32], on the importance of embodiment). There are more notable works in this direction, which all put emphasis on the sensorimotor aspect of meaning [33] [34] [35] [36] [37]. Recent works in deep learning are also exploring grounding based on learning of physical commonsense knowledge through interaction with environmental objects. See [38], for example.

One interesting question is, does the range of possible motor behavior somehow limit the degree of understanding? That is, can organisms with higher degree of freedom and richer repertoire of actions gain higher level of understanding? I believe this is true. For example, recall the orientation perception thought experiment in Section 2. If the visuomotor agent was only able to move horizontally or vertically, but not diagonally, it would never be able to figure out what the 45^o and 135^o light bulbs (the neurons) mean. Intelligence is generally associated with the brain size or brain/body ratio, but what may also be very important is how rich the behavioral repertoire of the animal is. For example, all the animals we consider to be intelligent have such flexibility in behavior: primates, elephants, dolphins, corvids, and even octopuses. An extension of this idea is, can an agent extend its behavioral repertoire? This is possible by learning new moves, but it is also possible by using tools. The degree of understanding can exponentially grow if the agent can also construct increasingly more complex tools. This I think is one of the keys to human's superior intelligence. See [39] for our latest work on tool construction and tool use in a simple neuroevolution agent, and our earlier work on tool use referenced within. On a related note, also consider the works by Zhao et al. [40], where they showed how novel robot morphology and controllers can be discovered through graph heuristic search. Although this is not strictly tool construction, the same methodology could be extended to tool construction and use.

In section 2, I proposed the internal state invariance criterion, within the context of reinforcement learning. This raises an interesting idea regarding rewards in reinforcement learning. In traditional reinforcement learning, the reward comes from the external environment. However, research in reinforcement learning started to explore the importance of rewards generated from within the learning agent. This is called "intrinsic motivation" [22] [41], and the internal state invariance criterion could be a good candidate for this. In this view, intrinsic motivation also seems to be an important

ingredient for meaning that is intrinsic to the learning system. Another related work in this direction is [42], based on the criterion of independently controllable features (also see [43] on “empowerment”). The main idea is to look for good internal representations where “good” is defined by whether an action can independently control these internal representations or not. With this, both the perceptual representations and the motor policy are learned. This kind of criterion can be internal to the agent, thus, keeping things intrinsic, while allowing the agent to understand the external environment. Also see [44] for our earlier work on co-development of visual receptive fields (perceptual representations) and the motor policy.

Next, I would like to discuss various mechanisms that can serve as memory, and how, in the end, they all lead to prediction. In neural networks, there are several ways to make the network responsive to input from the past. Delayed input line is one, which allows a reactive feedforward network to take input from the past into consideration when processing the current input (see e.g., [9]). Another approach is to include recurrent connections, connections that form a loop. More sophisticated methods exist such as Long Short Term Memory (LSTM), etc., but generally they all fall under the same banner of recurrent neural networks. Finally, there is a third category that can serve as a memory mechanism, which is to allow feed forward neural networks to drop and detect token-like objects in the environment (a form of stigmergy). We have shown that this strategy can be used in tasks that require memory, just using feed forward neural networks [45]. From an evolutionary point of view, reactive feedforward neural networks may have appeared first, and subsequently, delay and ability to utilize external materials may have evolved (note that this is different with systems that have an integrated external memory, e.g., Differentiable Neural Computers [46]). Further development or internalization of some of these methods (especially the external material interaction mechanism) may have led to a fully internalized memory. These memory mechanisms involve some kind of recurrent loop (perhaps except for the delayed input case), thus giving rise to dynamic internal state (see [45] for a more detailed discussion on this, in relation to olfaction and hippocampal function) . As we have seen in section 3, in such a system, networks with predictive dynamics have a fitness advantage, and thus such phenotypes will proliferate.

Continuing with the discussion on prediction, let us consider the ideas by Jun Tani. In his book [47] (pp. 161-172), he also talks about predictive dynamics and self-consciousness. In his work, prediction is mostly about the sensory consequence of action, and about the error in this prediction (also see O’Regan and Noë’s notion of sensorimotor contingencies [48] and related ideas in Hawkins [49] pp. 139). Tani argues that self-consciousness arises when this prediction error is high. This may be counter to my notion of high predictability correlating with consciousness. However, Tani’s

formulation and my formulation are not directly comparable, since in Tani's case, the prediction error is computed by directly comparing the incoming sensory stimuli and the predicted sensory stimuli, while in my case, prediction is based purely on the internal state (the hidden state). I think these two views may be complementary. Further research into this may be needed, and the outcomes are expected to be synergistic. Lastly, I would like to mention that some recent advances in machine learning are benefitting from the use of prediction as a learning objective/criterion. In machine learning situations where explicit target values are rare or task-specific rewards are very sparse, it is a challenge to train effectively the learning model. Recent work by Finn and Levine [7] (and others) showed that learning motor tasks in a completely self-supervised manner is possible without detailed rewards, by using a deep predictive model, which uses a large data set of robotic pushing experiment. This shows a concrete example where prediction can be helpful to the agent. See [7] for more references on related approaches that utilize prediction.

Finally, let us examine question asking. As briefly hinted in Section 4 (citing [17]), generating questions or posing problems can be viewed as generating new goals. Similar in spirit with Schmidhuber's Powerplay [19], Florensa et al. proposed an algorithm for automatic goal generation in a reinforcement learning setting [50] (also see the works by Stanley et al. [51]). The algorithm is used to generate a range of tasks that the agent can currently perform. A generator network is used to propose a new task to the agent, where the task is drawn from a parameterized subset of the state space. A significant finding based on this approach is that the agent can efficiently and automatically learn a large range of different tasks without much prior knowledge. There have since been several different works on this topic. In [52], Misra et al. proposed learning by asking questions, and applied it to a visual question answering domain for automated curriculum discovery, and Akakzia et al. showed the connection between grounding and autonomously acquiring skills through goal generation [53]. These results show the powerful role of question asking in learning agents.

6. Conclusion

In this chapter, I talked about meaning vs. information, prediction vs. memory, and question vs. answer. These ideas challenge our ingrained views of brain function and intelligence (information, memory, and problem solving), and we saw how the momentum is building up to support the alternative views. In summary, we should pay attention to (1) meaning, and how it can be recovered through action, (2) prediction as a central function of the brain and artificial intelligence agents, and (3) question asking (or problem posing) as an important requirement for robust artificial intelligence, and the need to understand question asking mechanisms in brain science. In all three cases, we

also learned that taking the “internal” perspective is important, and that this can lead to different perspectives that can give us new insights.

Acknowledgments
In this revised Chapter, most notably, figures were added for a better explanation of the concepts discussed in the text, the discussion section was greatly expanded to include relevant references that were omitted, and latest references that appeared in the meanwhile were added. I would like to thank Asim Roy and Robert Kozma, who, together with myself, chaired a panel discussion on “Cutting Edge Neural Networks Research” at the 30th anniversary International Joint Conference on Neural Networks in Anchorage, Alaska in 2017, where I had the opportunity to refine many of my previous ideas, especially on meaning vs. information. I would also like to thank the panelists Peter Erdi, Alex Graves, Henry Markram, Leonid Perlovsky, Jose Principe, and Hava Siegelmann, and those in the audience who participated in the discussion. The full transcript of the panel discussion is available at https://goo.gl/297j1d. Figure 5(c)-(d) were reproduced from [15] by permission from IEEE (license #5310531004085). Finally, I would like to thank my current and former students who helped develop and test the ideas in this chapter, and Takashi Yamauchi for his thoughtful feedback.

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal,* vol. 27, no. 3, pp. 379-423, 1948.

[2] J. R. Searle and others, "Minds, brains, and programs," 1980, pp. 201-224.

[3] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena,* vol. 42, no. 1-3, pp. 335-346, 1990.

[4] G. Buzsáki, The brain from inside out, Oxford University Press, 2019.

[5] Y. Choe and N. H. Smith, "Motion-Based Autonomous Grounding: Inferring External World Properties from Internal Sensory States Alone," in *ConferenceProceedings of the 21st National Conference on Artificial Intelligence(AAAI 2006)*, 2006.

[6] A. Clark, "Whatever next? Predictive brains, situated agents, and the future of cognitive science," *Behavioral and brain sciences,* vol. 36, no. 3, pp. 181-204, 2013.

[7] C. Finn and S. Levine, "Deep visual foresight for planning robot motion," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[8] S. J. Thorpe and M. Fabre-Thorpe, "Seeking categories in the brain," *Science,* vol. 291, no. 5502, pp. 260-263, 2001.

[9] K. Nguyen and Y. Choe, "Dynamic Control Using Feedforward Networks with Adaptive Delay and Facilitating Neural Dynamics," in *ConferenceProceedings of the International Joint Conference on Neural Networks*, 2017.

[10] H. Lim and Y. Choe, "Extrapolative Delay Compensation Through Facilitating Synapses and Its Relation to the Flash-lag Effect," *IEEE Transactions on Neural Networks,* vol. 19, pp. 1678-1688, 2008.

[11] F. Mormann and C. Koch, "Neural correlates of consciousness," *Scholarpedia,* vol. 2, no. 12, p. 1740, 2007.

[12] Y. Choe, J. Kwon and J. R. Chung, "Time, Consciousness, and Mind Uploading," *International Journal on Machine Consciousness,* vol. 4, pp. 257-274, 2012.

[13] N. Humphrey, A History of the Mind, New York: HarperCollins, 1992.

[14] J. Yoo, J. Kwon and Y. Choe, "Predictable Internal Brain Dynamics in EEG and Its Relation to Conscious States," *Frontiers in Neurorobotics,* vol. 8(00018), 2014.

[15] J. Kwon and Y. Choe, "Internal State Predictability as an Evolutionary Precursor of Self-Awareness and Agency," in *ConferenceProceedings of the Seventh International Conference on Development and Learning*, 2008.

[16] G. Claxton, Hare Brain, Tortoise Mind: Why Intelligence Increases When You Think Less, Hopewell, NJ: The Ecco Press, 1999.

[17] Y. Choe and T. A. Mann, "From Problem Solving to Problem Posing," *Brain-Mind Magazine,* vol. 1, pp. 7-8, 2012.

[18] E. A. Silver, "On mathematical problem posing," *For the learning of mathematics,* vol. 14, no. 1, pp. 19-28, 1994.

[19] J. Schmidhuber, "POWERPLAY: Training an Increasingly General Problem Solver by Continually Searching for the Simplest Still Unsolvable Problem," *Frontiers in Psychology,* vol. 4, no. 313, 2013.

[20] M. Cakmak and A. L. Thomaz, "Designing Robot Learners that Ask Good Questions," in *ConferenceProceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction*, 2012.

[21] F. Liu, T. Xiang, T. M. Hospedales, W. Yang and C. Sun, "iVQA: Inverse visual question answering," in *ConferenceProceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[22] S. Forestier, R. Portelas, Y. Mollard and P.-Y. Oudeyer, "Intrinsically motivated goal exploration processes with automatic curriculum learning," *arXiv preprint arXiv:1708.02190,* 2017.

[23] C. Colas, T. Karch, N. Lair, J.-M. Dussoux, C. Moulin-Frier, P. F. Dominey and P.-Y. Oudeyer, "Language as a cognitive tool to imagine goals in curiosity-driven exploration," *arXiv preprint arXiv:2002.09253,* 2020.

[24] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," *arXiv preprint arXiv:1310.4546,* 2013.

[25] B. M. Lake and G. L. Murphy, "Word meaning in minds and machines," *arXiv preprint arXiv:2008.01766,* 2020.

[26] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell and others, "Language models are few-shot learners," *arXiv preprint arXiv:2005.14165,* 2020.

[27] L. Wittgenstein, Philosophical Investigations, Oxford: Blackwell, 1953.

[28] B. Baker, B. Lansdell and K. Kording, "A philosophical understanding of representation in neuroscience," *arXiv preprint,* p. arXiv:2102.06592, 2021.

[29] A. M. Glenberg and D. A. Robertson, "Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning," *Journal of memory and language,* vol. 43, no. 3, pp. 379-401, 2000.

[30] M. Mitchell, "Why AI is Harder Than We Think," *arXiv preprint arXiv:2104.12871,* 2021.

[31] F. J. Varela, E. Thompson and E. Rosch, The Embodied Mind: Cognitive Science and Human Experience, Cambridge, MA: MIT Press, 1991.

[32] H. J. Chiel and R. D. Beer, "The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment," *Trends in neurosciences,* vol. 20, no. 12, pp. 553-557, 1997.

[33] R. A. Téllez and C. Angulo, "Acquisition of meaning through distributed robot control," in *ConferenceProceedings of the ICRA Workshop on Semantic information in robotics*, 2007.

[34] A. Laflaquière, J. K. O'Regan, B. Gas and A. Terekhov, "Discovering space—Grounding spatial topology and metric regularity in a naive agent's sensorimotor experience," *Neural Networks,* vol. 105, pp. 371-392, 2018.

[35] A. K. Engel, A. Maye, M. Kurthen and P. König, "Where's the action? The pragmatic turn in cognitive science," *Trends in cognitive sciences,* vol. 17, no. 5, pp. 202-209, 2013.

[36] J. Modayil and B. Kuipers, "Autonomous development of a grounded object ontology by a learning robot," in *ConferenceProceedings of the national conference on Artificial intelligence*, 2007.

[37] J. M. Stober, "Sensorimotor embedding: a developmental approach to learning geometry," 2015.

[38] R. Zellers, A. Holtzman, M. Peters, R. Mottaghi, A. Kembhavi, A. Farhadi and Y. Choi, "PIGLeT: Language Grounding Through Neuro-Symbolic Interaction in a 3D World," *arXiv preprint arXiv:2106.00188,* 2021.

[39] R. Reams and Y. Choe, "Emergence of Tool Construction in an Articulated Limb Controlled by Evolved Neural Circuits," in *ConferenceProceedings of the International Joint Conference on Neural Networks*, 2017.

[40] A. Zhao, J. Xu, M. Konaković-Luković, J. Hughes, A. Spielberg, D. Rus and W. Matusik, "RoboGrammar: graph grammar for terrain-optimized robot design," *ACM Transactions on Graphics (TOG),* vol. 39, no. 6, pp. 1-16, 2020.

[41] A. G. Barto, S. Singh and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills," in *ConferenceProceedings of the 3rd International Conference on Development and Learning*, 2004.

[42] V. Thomas, J. Pondard, E. Bengio, M. Sarfati, P. Beaudoin, M.-J. Meurs, J. Pineau, D. Precup and Y. Bengio, "Independently controllable features," *arXiv preprint arXiv:1708.01289,* 2017.

[43] N. C. Volpi and D. Polani, "Goal-directed Empowerment: combining Intrinsic Motivation and Task-oriented Behaviour," *IEEE Transactions on Cognitive and Developmental Systems,* 2020.

[44] H.-F. Yang and Y. Choe, "Co-development of visual receptive fields and their motor-primitive-based decoding scheme," in *ConferenceProceedings of the International Joint Conference on Neural Networks 2007 Post conference Workshop on Biologically-inspired Computational Vision (BCV) 2007*, 2007.

[45] J. R. Chung and Y. Choe, "Emergence of Memory in Reactive Agents Equipped with Environmental Markers," *IEEE Transactions on Autonomous Mental Development,* vol. 3, pp. 257-271, 2011.

[46] A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. G. Colmenarejo, E. Grefenstette, T. Ramalho, J. Agapiou and others, "Hybrid computing using a neural network with dynamic external memory," *Nature,* vol. 538, no. 7626, pp. 471-476, 2016.

[47] J. Tani, Exploring robotic minds: actions, symbols, and consciousness as self-organizing dynamic phenomena, Oxford University Press, 2016.

[48] J. K. O'Regan and A. Noë, "A Sensorimotor Account of Vision and Visual Consciousness," *Behavioral and Brain Sciences,* vol. 24, pp. 939-973, 2001.

[49] J. Hawkins, A thousand brains: A new theory of intelligence, New York, NY: Basic Books, 2021.

[50] C. Florensa, D. Held, X. Geng and P. Abbeel, "Automatic goal generation for reinforcement learning agents," in *International conference on machine learning*, 2018.

[51] R. Wang, J. Lehman, J. Clune and K. O. Stanley, "Paired open-ended trailblazer (POET): Endlessly generating increasingly complex and diverse learning environments and their solutions," *arXiv preprint arXiv:1901.01753,* 2019.

[52] I. Misra, R. Girshick, R. Fergus, M. Hebert, A. Gupta and L. Van Der Maaten, "Learning by asking questions," in *ConferenceProceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[53] A. Akakzia, C. Colas, P.-Y. Oudeyer, M. Chetouani and O. Sigaud, "Grounding Language to Autonomously-Acquired Skills via Goal Generation," *arXiv preprint arXiv:2006.07185,* 2020.