Perception and time in Artificial General Intelligence.

Andrey Makushkin
toandrey@yahoo.com
August 5, 2019

Summary:
Intelligence is an emergent behavior attributed to entities possessing mechanisms such as perception, short and long term memory, ability to learn or forget, attention, cognition, reasoning, prediction, reflexes, embodiment, actuation and many other. Perception is considered the most important mechanism by many researchers.[1][2]

In traditional computing, data is received or acquired and transformed by computers. The output of these computations is perceived and interpreted by humans. When building Artificial General Intelligence (AGI) our goal is to move this perception boundary inside the system.

The goal of this paper is to restrict the definition of perception in the discipline of AGI to the detection mechanism. This paper tries to explain this statement through examples and elaborate on the implications. Another goal is to underline the importance of time in AGI systems.

Introduction:

In the fields of AI and AGI, currently accepted definition of perception can be summarized as "perception is a change in internal state as a result of external stimuli". Better descriptions take time into account usually by including something abstract, for example, temporal patterns of neural activity. These definitions do not adequately describe the mechanism.

Inadequacies of the existing definitions are exposed by the following example. A rock's internal state can be changed by heating it. It will have a short term memory of it because it remains hot for some time. It can have a long term memory of it if that changes its crystalline structure. Geologists make careers out of studying long term memories of the rocks. The rock can reflect or absorb light, conduct or insulate electric currents. It can also actuate by falling as a result of changing its state. Taking all that complex behavior into account not many will claim that the rock can perceive its environment. Drawing this analogy, CCD and CMOS sensors which are designed to sense the environment and whose wafers are made from silica are not able to perceive the environment. On the other hand, many will agree that neurons, biological sensors, are able to perceive their environment. Therefore a different process takes place inside a neuron. Just like in the rock, neuron's state (its membrane potential) is changed by external stimuli. External stimuli constitute sensory stimuli or synaptic potential. When the membrane potential reaches a certain voltage level, it is detected by the neuron and the neuron fires. Neuron's detection mechanism is significantly different from the processes occurring in the rock.

This brings us to the following definition. Perception is a low level mechanism that can be composed to form higher level functions. It should not be confused with the interpretation mechanism, cognition. Perception is an act of detection of an internal state change described in terms of time. Detection is described by when it has occurred as opposed to how the internal state changed. In other words, detection of the state change inside the system should not be described by further changing the system state. This is similar to neurologists modeling neural activity as one dimensional point processes with time labeling the axis vs. sampling the membrane potentials.[D4] Any other way of observing the environment results in a symbolic world representation.

Perception should be modeled as a set of interacting point processes. It is the author's goal to show the reasons for doing so and the advantages by describing the detection mechanism and its relationship with time in view of different disciplines. Some of the arguments found here will use an idea of a non-conscious observer borrowed from the information theory.[D1] One deviation from the definition of and observer is that it has a dual nature. It is a process and an object at the same time. Detection occurs within this process. In the example above, processes in the rock were started externally whereas an observer has an internal running process. It will also rely on the notion of an object's state - the information that describes the object. The author will sometimes refer to an observer's own internal state. Information that describes the observer and most importantly how that information is changed by its environment.

Measure, values, numbers, symbols, state, perception and philosophy:
When we are looking at observations expressed in terms of values we are frequently looking at the number of events per unit of time. A count of how many times a measuring stick fits. How many changes have been detected. Measuring the state of the environment produces values such as length, angles or voltage levels, frequently expressed as numbers in digital systems. Numbers are measures of an object's state (or rate of state change) from an absolute reference point such as zero. The reference point and scale usually have nothing to do with the object whose properties are measured. They are qualities of an observer. I measure in inches, you measure in centimeters. I measure in degrees, you measure in radians. Does that thermometer measure the temperature in Fahrenheit or Celsius? What is the difference between 20 and 68? None, because 20 degrees Celsius is 68 degrees Fahrenheit. Is five longer than five? Yes, because five inches are longer than five centimeters. Values are meaningless without units. If the units do not match, things cannot be compared.

This leads to a conclusion that although a measure might have a meaning for a single observer, it is meaningless to another observer unless units are agreed upon. Ironically quantification, a concept that has to give meaning to observation is the reason observations lose their grounding.[D2][D3] Quantification forms a symbolic world representation. It would be easier to understand perception through a mechanism where the environment changes an observer's state without agreeing on units than transferring information in terms of numbers or symbols. The observer simply adapts to how its state changes.

Before proceeding, let me reference some historical views on perception. The first well-known models of perception are Plato's theory of forms and his "Allegory of the Cave". Plato describes objects being perceived as shadows without an observer ever knowing their true form. He also adds that a true philosopher thinks in terms of forms.

Numbers are projections of real-world phenomena onto an axis. They are Plato's shadows. Every time a measurement is made, a shadow is produced. The alternatives are Plato's forms. Forms cannot be measured, described as a symbol, or transmitted. Forms can only be detected.

This closely relates to the problem of qualia which cannot be expressed symbolically. Qualia is a process of perception of a form. Biological neurons detect Plato's forms.

Any observer cannot directly measure the state of some object such as its color or temperature. It can only observe changes in the environment. The object of observation does not always need to change. A change in the medium is sufficient to deliver the information to the observer. A photon arriving at the retina, hair deflecting in a cochlea, tactile input, molecule locking into a receptor in the nose, they are all events occurring in the sensor's environment.

Whenever a change occurs it has to be described in terms of time. For example, a drop in temperature freezes the lake, not the cold itself. The fact that it's cold can tell us about the state of the lake being frozen. It does not tell us how long. Only the fact that the temperature is changing can describe that the state of the lake is changing.

Meaning of the information changes over time. Information that there is a wall ten feet ahead might be meaningless to a self driving car a year after because by then it could have been involved in an accident and recycled into a toaster. If information is represented by a symbol, a millisecond later a different symbol has to be used because the information has already changed. This is one more reason to express information in terms of time.

Time instance is the smallest quanta of information yet since time is continuous its resolution is limited only by an observer's precision. Detecting a change is much more precise in time than sampling. While sampling, the change could occur anywhere within the sample time interval.

There is another side effect of using a detection mechanism. Detecting a change will result in low jitter and therefore a constant delay while it propagates through the system from an observer to another observer. Similar to the propagation of the information about an event through the environment. These delays can be exploited by some mechanisms, for example, when locating a source of sound. Another example is mercury delay lines which were used in the early days of electronics to create computer memory.

There is one more advantage of using discrete perception related to the notion of causality. Imagine you have two containers A and B. Container A is being filled by two streams of liquid. Two more sources of liquid are dripping into container B. When container A overflows it is impossible to tell which stream caused it to overflow. When container B overflows, the source of the last drop can be determined.

Perception in Artificial Intelligence (AI):

First memorable mention of perception in AI has to be attributed to Frank Rosenblatt. In his 1957 paper "The Perceptron A Perceiving And Recognizing Automaton" he puts perception first in the list of "human-like functions". He also references Plato's' forms: "the form of a man seen from any angle". This reference has been greatly overlooked. In addition, Rosenblatt emphasizes the importance of time by dividing perceptrons into two categories: "momentary stimulus perceptrons and temporal pattern perceptrons".

Whereas Rosenblatt wanted his perceptrons to detect the forms, nowadays we have a more generic word for mechanisms of perception to detect forms. They are called features. A feature is a mechanism of perception. Features can be detected by current computation models and even used in a generative mode to create pictures. Features can be used in a generative mode because they embed information into its definition. Feature's importance as a mechanism of perception that can make real-world observations in terms of time is not well-known.

Are features the answer to perception? It depends on your point of view. In many current models feature detection or absence is represented by symbols 1 and 0. Some researchers think about these models of ANNs as symbolic in nature where symbols 0 or 1 are used to "communicate" among neurons.
The problem with this point of view is the "Symbol Grounding Problem".[D3]

"Symbol Grounding Problem" should not be just a red flag for symbolic AI but a red brick wall for any approach that uses state in its computation model. Zeros and Ones used in ANNs are not symbols. They represent time. There are problems with this way of representing time. Some of the problems are described in "Finding Structure In Time" paper by Jeffrey L. Elman. The important thing is to agree they represent time. Since no symbol transmission takes place, connectionism avoids the "Symbol Grounding Problem" inherent in symbolic systems.

However things are more complex than that. One and zero could represent an interval of time or a point in time. If you accept the above definition of perception, then ones and zeros represent points in time. Points in time at which change was detected. This has implications that widen the boundary between computationalism and connectionism. One of the arguments used to show equivalence of the two theories for the past seventy years was that "neural events and the relations among them can be treated by means of propositional logic".[7] Whereas one could argue that overlapping intervals of time could be connected via propositional logic operators, points in time require temporal logic. Since observers are not synchronized in time, probability that events are detected at exactly the same time tends to zero. Since the points can be a millisecond or a million years apart, temporal logic is required to reason about them.

Another ANN related fact that underlines the importance of the time aspect of the computational models is the success of LSTM algorithm. Its success can be attributed to a clever timing mechanism that makes it stand out among other recurrent ANNs. The mechanism facilitates processing of stimuli from time intervals far apart.

If you train a supervised learning algorithm to give you a prediction one day into the future, it will not be able to predict what happens in an hour. This fact is counterintuitive. At first, it might appear that making an algorithm an online learning algorithm might solve the problem. However, this is not the case and the problem should be looked at from a different angle. There are two types of AI. A static world, turn-based and a dynamic world AI. In a dynamic world, processes are observed. Time is a fundamental description of the perceived state of the process (signals). In static world AI, a system is presented with data without a time component. Such is done in turn-based games, pattern recognition, etc. Static world AI can be tricked to perform on dynamic world problems by saying take turn every few milliseconds or run every time a sensor is read. This approach can even be used in robotics. However, it impedes research progress by climbing local minima.

Neurology:

A hypothesis that brain is required for mobility is given to us courtesy of the sea squirt. It turns out every voluntary answer your brain gives to any question involves muscle movements.[3] We can look at a muscle fiber as an interface between a brain and the outside world. A muscle fiber only needs to know if it has to contract right now. The only question that the brain has to answer then becomes "should I twitch that muscle fiber right now?" The answer is always "yes" meaning "contract right now" or it does not arrive at all. "Now" is the best answer! All information available before right now is taken into account. Answering by providing a specific time or a condition that has to occur in the future runs a risk that new information will become available and a decision will no longer be optimal. This proves the importance of perception in terms of time.

In biological NNs over two hundred neurotransmitters are found. If an analogy with ANN is drawn, over two hundred different messages or symbols can be sent or received by neurons. Current theories do not address this variety of messages processed by biological neurons. In neurology, mechanisms of changing the internal state of a neuron are many and variable. Neuron's membrane potential can be changed by photons, chemical, electrical or mechanical means. Frequently indirectly through the flow of sodium and potassium ions across a membrane. Mechanisms can be inhibiting or exciting. These mechanisms are important but are not common to all types of cells. The common principle is the detection of internal state change. How detection works also varies. Adaptation indicates detection thresholds are not constant.[D5] On the other hand, pain neuronal pathways are not learned and pain receptors do not become less sensitive over short time.[5]

Once the mechanism of perception is accepted, one can argue about the mechanisms of state change. Whether they require inhibition. If synapses should be modeled as weights or oscillators. It might be discovered that mechanisms of stimuli are not that important. What is important is that change in the internal state of the neuron is detected. This detection is represented in the system in terms of time by when it was detected. As a result, functionality of sensory neurons can be treated the same as cortical or peripheral neurons. They are all change detectors. In ANNs this would translate to input and hidden layer neurons having the same detection mechanism. Analogous, in spiking NNs, spikes are not important. They are simply representations of the detector's output.

Why detection? Why does the observer need to detect its own state change? What about other mechanisms? Can an observer after its state has been modified go on to continuously modify another observer's state proportionally to its own state change? Similar to analog electronics where for many components changing its input changes its output continuously. This approach is used in analog computing which is also not symbolic in nature. However, this approach leads to a continuous expansion of energy. The detection mechanism is also a biological approach to minimizing energy consumption. According to Rashevsky, "for a set of prescribed biological functions an organism has the optimal possible design with respect to economy of material used, and energy expenditure, needed for the performance of the prescribed function." [4]

It takes energy to change state. This energy should not be expanded unless the observer's state in the previous example has changed. Otherwise, by expanding this energy, the observer changes its own state nondeterministically. However, a system needs to react to external stimuli and not just its unit's random internal state changes. In addition, the observer needs to reset its state in such a way that when its state is changed again, the net change can be detected.

On the importance of time in physics and engineering:
In classical physics, one wants to describe the world in terms of functions. For example, in an experiment where a ball is moving left and right on a track the ball's position can be described as a function of time f(t). This description breaks down when something unaccounted for happens. For example, the ball is picked up and later placed back on the track. This action can be defined as a discontinuity in f(t). The discontinuity can be described in terms of the time interval because math describes the world in an instance of time as a function of time.

In Engineering non-linear processes are often modeled by linear functions on some intervals. The function can be chosen by an engineer or fitted by linear regression. When linear approximation is used for a function of time, a time interval during which it is valid has to be defined. This is similar to subsumption architecture in robotics where state machines describing non-linear behavior become valid during some periods of time. Describing everything in a system (perception/ internal representations /actions) in terms of time fits well with the techniques described above.

Probability, statistics and Digital Signal Processing (DSP):
Sampling a random process corresponding to a real-world phenomenon at an arbitrary frequency does not guarantee exceeding Nyquist rate.[D8][D9] Sampling at the maximum sensor frequency produces excessive amounts of data. For example, a sensor sampling brightness 30 times per second produces 2,592,000 samples in 24 hours. This is excessive if the goal is to figure out whether it is night or day. Even if data is not retained, it takes energy to process it.

An alternative is to detect changes in the signal and record the amplitude and time of the change. As the speed of an analog to digital converter (ADC) increases, amplitude changes from previous value take on a set {-1,0,1} due to the limits of sensor resolution. The sensor becomes a change detector. This fact is exploited in the "One-bit Audio" encoding.

Observing the signal via multiple observers that detect changes instead of sampling is a better model that uses the mechanism of perception. For example, observer one detects a change in one unit from an absolute. Observer two detects a change in two units and so on. When noise is filtered out, the approach of detecting the changes produces the optimal rate for capturing the signal. Detecting signal changes guarantees to capture the information needed to reproduce the signal. Detection mechanism does not suffer from sampling below Nyquist rate.

In DSP, random variables (random processes) such as phase, amplitude and frequency are sometimes defined a priori. Transmitter manipulates these variables creating a signal. Frequency is a known random variable when designing filters. It is easy to fall into a trap thinking that using these random variables in any signal will provide some insight into the information content. The same applies to using statistical properties of the signal such as mean, median, deviation, distribution to gain insight. The problem is that any complex signal is a composition of multiple random variables. Without knowing all of them, it is impossible to understand how one of them contributes to the statistics of a complex signal.

There is an alternative approach of creating statistical experiments to extract information. The problem is, how does one construct an experiment automatically? How does one go from observing a continuous random process to an event? Sampling cannot be used to produce experiment outcomes. As we know from Bertrand paradox, method of selection influences the distribution of the random variable.[D6] While sampling, changing the sample rate changes the statistics. There is no single way to create an experiment by sampling. The mechanism of perception however, can be used to define and run statistical experiments. Detecting a change in a signal produces an experiment outcome. Each possible outcome is defined by the presence of an observer. Since observers have different properties, each observer in a system observing the same process detects distinct outcomes at different times. Similar observers change or eliminate each other as seen in some biological mechanisms.[6] The concept of feature binding (BP2) helps explain how the boundaries of the experiment are defined.[D7] Since their observer's properties are different, every system conducts different experiments with the same real world phenomenon. For example, dogs cannot see colors whereas we can.

At this point the importance of an ability to actuate and embodiment has to be mentioned. It enables modification of statistical experiments. For example, a coin is laying on the road. Cars are passing over it and push the coin around but it does not flip. It always lays on one side. Observing this experiment without an ability to modify it will not tell you if this is a fair coin.

When working with data, researchers often face combinatorial explosion. Having a time component may help reduce the number of possible combinations of random variable realizations.

Data may exhibit a correlation without causality. Causality on the other side will express itself as a correlation in data from different time intervals. To avoid an inverse error one needs to specify that correlation of data from different time intervals does not imply causality. In order to reason about causality, one needs to establish the notion of time in the reasoning framework. This is easier done when information is expressed in terms of time as for example in a one dimensional point process with time labeling the axis.

Conclusion:

Contemporary directions of research aligned with this view on perception already include neuromorphic computing, spiking ANNs, BEAM robotics, study of coupled oscillators and study of point processes. It's important to revisit and reconsider classic research with this new outlook on perception to find what was missed and correct the assumptions.

Glossary:

[D1] An Observer is a system that receives information about an object.
[D2] Quantification: https://en.wikipedia.org/wiki/Quantification_(science)
[D3] Symbol grounding: https://en.wikipedia.org/wiki/Symbol_grounding_problem
[D4] Point process: https://en.wikipedia.org/wiki/Point_process
[D5] Adaptation: https://en.wikipedia.org/wiki/Neural_adaptation
[D6] Bertrand paradox: https://en.wikipedia.org/wiki/Bertrand_paradox_(probability)
[D7] Feature binding: https://en.wikipedia.org/wiki/Binding_problem
[D8] Random process (stochastic process):
https://en.wikipedia.org/wiki/Stochastic_process
[D9] Nyquist rate: https://en.wikipedia.org/wiki/Nyquist_rate

References:

[1] Frank Rosenblatt, (1957). The Perceptron A Perceiving and Recognizing Automaton (Project Para). Report No. 85-460-1. Cornell Aeronautical laboratory, inc. Buffalo, N.Y.
[2] Stephen Wolfram, (2002). A New Kind of Science, chapter 10 "Perception", Wolfram Media. ISBN:1579550088.
[3] https://www.ted.com/talks/daniel_wolpert_the_real_reason_for_brains
[4] Nicolas Rashevsky, (1960). Mathematical Biophysics: Physico-Mathematical Foundations of Biology, vol. 2, p. 292
[5] http://www.newworldencyclopedia.org/entry/Pain
[6] M. Favero, G. Busetto, A. Cangiano (2012). "Spike timing plays a key role in synapse elimination at the neuromuscular junction"
PNAS, DOI 10.1073/pnas.1201147109
[7] W.S. McCulloch, W. Pitts, (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics, Vol. 5 pp. 115-133