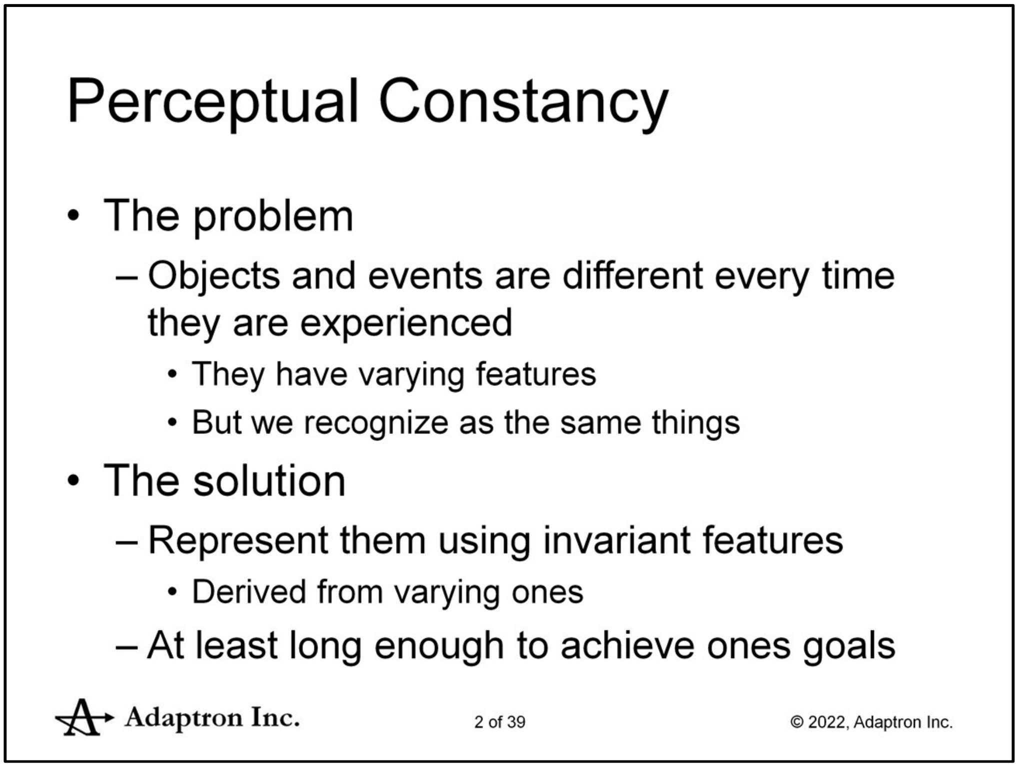
Grounding Invariant Features with Binons for Perceptual Constancy

Brett Martensen
Adaptron Inc.

24th April 2022

I'm planning on writing a paper based on this presentation and submitting it to the NeSy 2022 workshop.
It is one of a number of workshops in the Second International Conference on Learning and Reasoning (IJCLR 2022) in Windsor, UK on 28th – 30th September. Papers are due by May 31st.



When it comes to perception, we think we have invariant symbolic mental representations of the things that we experience even though they are perceptually different every time they occur. Our experiences are continuously changing and never occur exactly the same way twice.

So how do we obtain invariant representations of objects and events from their varying features?

And then how do we recognize them when they re-occur? Obviously we need to derive invariant features from the varying ones. And our mental representations have to be constant long enough that we can achieve our goals.

Features

- Features
  - Three types
    - Properties (e.g. loudness, size, duration)
    - Relational (e.g. on, above, left of, before, after)
    - Parts (e.g. handles on doors, singing at a party)
  - Dimensions
    - Spatial (e.g. size, above and handles)
    - Temporal (e.g. duration, before and singing events)

Adaptron Inc. 3 of 39

We need to be clear on what we mean when we use words like features.
The three types are property features, relational features and part features. They can all be spatial or temporal.
For example the size of a door is a numeric property and it is spatial.
The duration of a birthday party is a numeric property of an event and events are temporal.
A handle on a door is a spatial relational feature while blowing out the candles on the birthday cake before eating it is a temporal relationship.
Door handles are objects and they are parts of doors, which is spatial.
A song being sung at a birthday party is an event; it is part of the party which takes time to happen and therefore is temporal.

Categories

- Invariant features are grouped to represent categories [1]
  - Category members share common features
- Differences are invariant
  - Distance of 7 independent of positions
    - 1 → 8, 6 → 13, 28 → 35 etc.
- Ratios are invariant
  - Ratio 1/2 independent of source values
    - 2/4, 3/6, 4/8, 5/10 etc.
- Differences and ratios are simple categories

Adaptron Inc. 4 of 39

How do we produce single representations for categories such as doors and birthday parties?
We need to build it out of invariant features. But how do we create invariant features from varying ones?
Differences between numbers are invariant with respect to the varying numbers used to calculate them.
For example the distance of 7 is independent of the number combination used in its calculation.
But the distance can still vary. What about ratios such as ½? They also are independent of the pair of numbers used in their calculation but the ratio is a number and can vary.
But any given difference or ratio does represent a category of a pair of numbers.
Categories of things group their members based on some common features, so differences and ratios are a good place to start when creating invariant features.

Perception

- Process to:
  - Sense, recognize and encode things
  - Grounded on the senses
- Produces percepts
  - Mental or artificial representations of objects and events
- Forms deep compositional hierarchies [2]

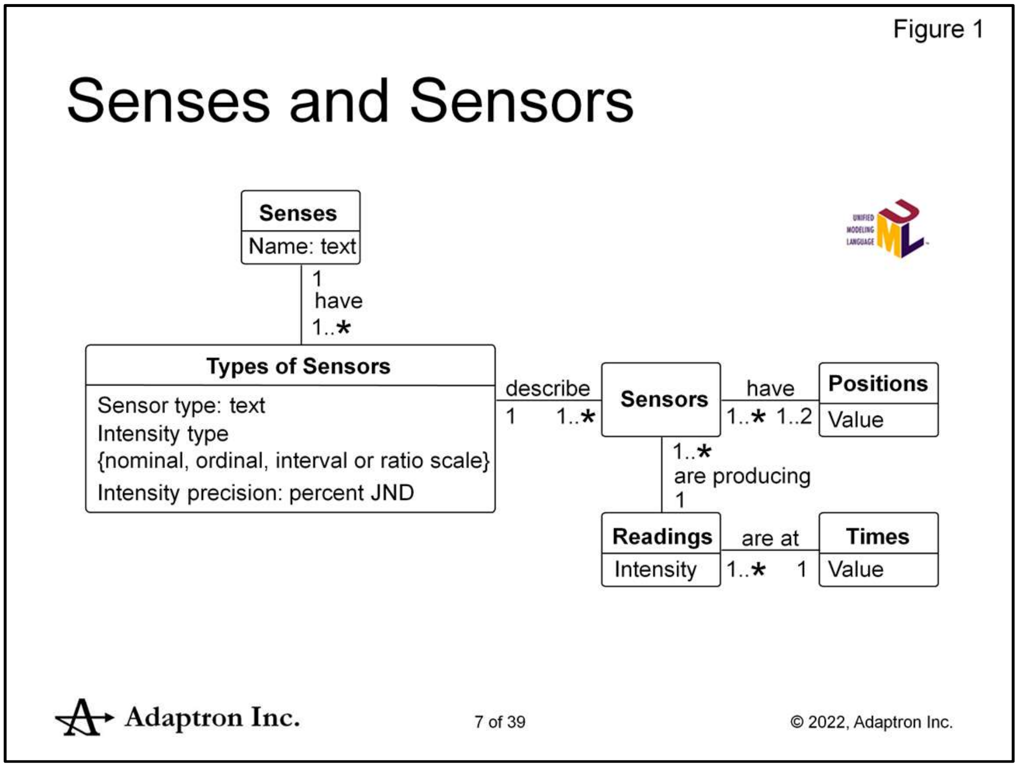
Perception is a process by which we sense, recognize and encode memories of things. Based on the sense involved, these things could be visual objects, sounds, pressures on the skin, tastes or smells. This means they are grounded on the senses.
Sensing is the detection of these things, recognizing is identifying them when they are re-experienced and encoding is representing and memorizing them.
Their mental representations are called percepts.
Percepts are combined into deep hierarchies in which more complex things such as doors are composed of many simpler things such as handles, hinges and door slabs. And handles are composed of the latch, knobs and internal parts. That is a spatial example. A temporal one is a birthday party. It is made up of singing happy birthday, eating cake, opening presents and playing games.

Perception – Sensing

- Symbol Grounding Problem [3]
- Percepts get their meaning
  - Relative to the environment
  - Via the senses
- Need a general–purpose design of senses for general–purpose artificial agents

Adaptron Inc. 6 of 39

But there is a problem in perception called the Symbol Grounding Problem.
First, it assumes that the things we perceive and the concepts we think about are represented symbolically. This idea is reinforced by the fact that we use words to talk about them.
The problem is, how do these percepts and concepts get their meaning? What are they grounded or based on?
The answer is they are derived from what the senses can detect and measure.
So it seems reasonable if we want to build a general-purpose intelligent agent, we better have a general-purpose description of the senses.



This Unified Modeling Language (UML) class diagram is such a design.
It says that each sense has one or more types of sensors. For example the eye has cone cells for measuring colour intensity and rod cells detect brightness.
There are many sensors of each type. In the retina the cone and rod cells are arranged into sophisticated arrays with the densest concentration in the middle, the fovea.
If we consider a simpler two dimension array then each sensor has two positions, its x and y coordinates.
At any instance in time each sensor is producing a reading value which is the intensity of the type of property that it measures.
And therefore at any instance in time there will be many readings and the intensity of a reading may be coming from many sensors.
Explanations for the intensity types and their precision are coming up.

Intensity Types

- Non-symbolic values (numeric)
  - Interval scale (e.g. position, time)
  - Ratio scale (i.e. intensities)
    - E.g. lightness, pressure, loudness, weight
- Symbolic values (discrete)
  - Nominal (e.g. pencil, blue, door, ♫, ₿, @, Ω)
  - Ordinal (e.g. A, B, C, D, E, …)

The values measured via the sensors can be non-symbolic or symbol. Non-symbolic ones are numeric. Symbolic ones are discrete. They can be divided into interval and ratio scale. Interval scale values are numeric with an equal distance between the values, such as temperatures (in Celsius or Fahrenheit but not Kelvin). Positions or times and dates are also interval scale properties. In an interval scale, the value of zero is totally arbitrary.

Ratio scale values are real numbers and have a zero corresponding to nothing being measured. Examples are intensity values such as for lightness, pressure or loudness.

When you subtract one interval scale value from another you produce a ratio scale value. For example the difference between two positions is a distance.

When you divide one ratio scale value by another you produce a ratio. For example the ratio of two widths may indicate that one object is half as wide as another. You can do this with temperatures in degrees Kelvin but not Celsius or Fahrenheit.

Symbolic values are most useful in recognition because they are discrete and do not vary like numeric ones. No arithmetic can be performed on symbolic values.

They can be divided into nominal and ordinal scale.

Nominal scale values allow names (or integers) to be assigned to things. Words are typical examples of nominal scale symbolic values.

But they can also be individual symbols or characters.

An example of an ordinal scale would be the alphabet. The letters are in a particular order and can be sorted.

Sensors that produce numeric values are simple devices while those that produce symbolic values are far more sophisticated because they have to be intelligent enough to recognize something.

Numeric Precision

- Resolution of the sensors
  - determines the precision for non-symbolic values
- Weber's Law
  - Just-Noticeable Difference (JND)
  - Precision is a percentage of the value
    - 8% for weights
    - 10% for loudness
    - 14% for skin pressure

The precision to which an intensity value can be measured is dependent on the resolution of the sensors.
Ernst Weber determined how different two weights needed to be so their weights could be distinguished. The minimum required difference is known as the Just-Noticeable Difference (JND).
He determined that for human sensors, the resolution was different for each sense and that the resolution was dependent on the value. For example we need a 10% difference in loudness to notice a difference and that applies for a large range of values. This is also an effective way of cutting out low levels of noise.

Sensor Properties

- Core properties obtained from sensors
  - Position, intensity, time
  - Kept in a sensory memory
- Sensory memory layout
  - Same as the senses and their sensors

Adaptron Inc. 10 of 39

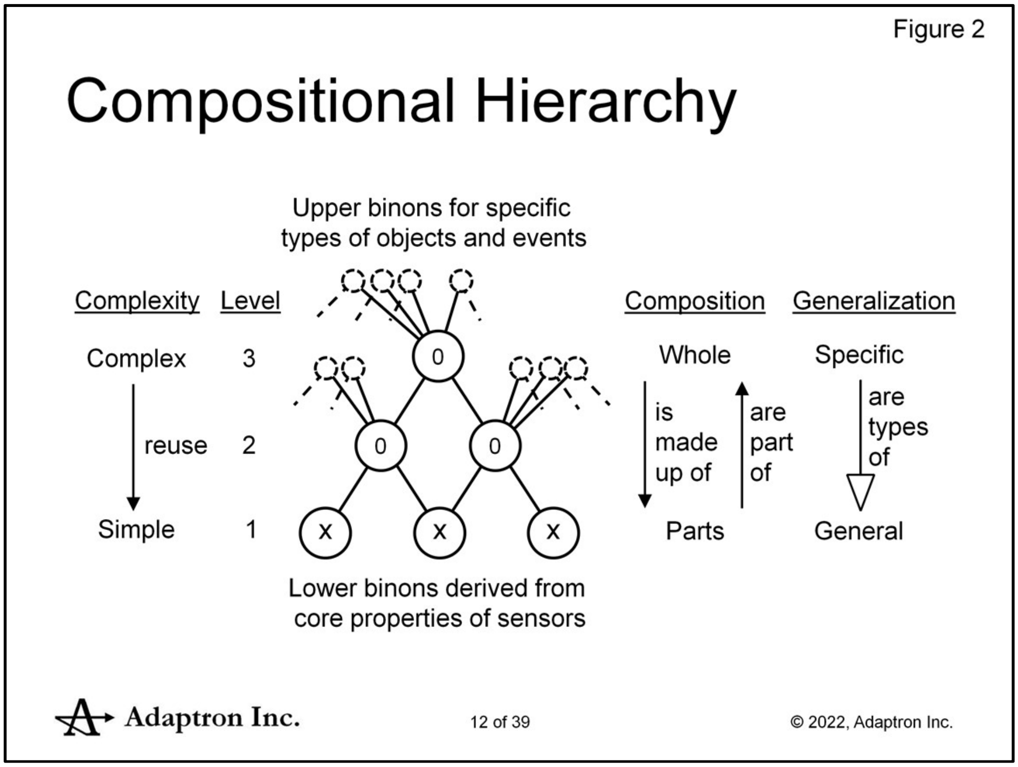
So the core sensor properties are position, intensity and time.
They are kept briefly in a sensory memory until the next time interval occurs.
The layout of the sensory memory is the same as that of the senses and their sensors.

Binons (Binary Neurons) [4]

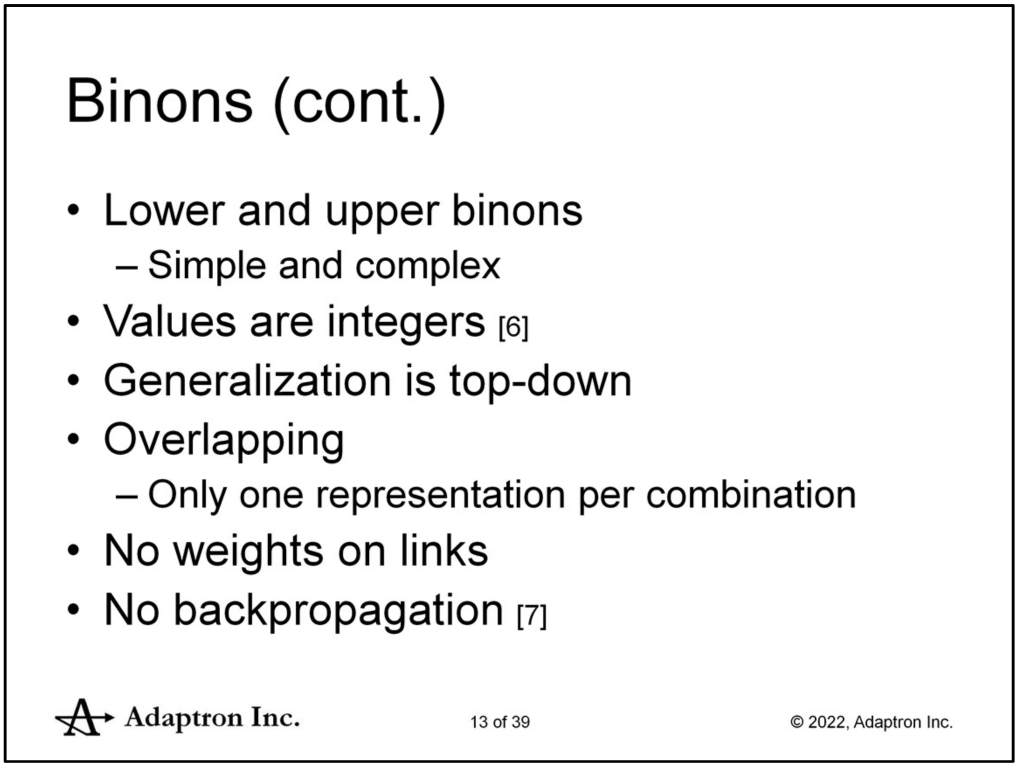
- General–purpose representations
  - For use in artificial intelligent agents
  - Represent percepts, actions, concepts and relationships between them
  - In perception they represent properties, objects and events
- Combined to form
  - Artificial neural networks (ANNs)
  - Deep overlapping compositional hierarchy [5]

Adaptron Inc. 11 of 39

Now lets describe binons (binary neurons).
After this introduction to binons I will come back to how the sensor properties are encoded into binons.
Binons are a general-purpose mechanism for representing percepts, actions and concepts in artificial intelligent agents.
In perception they are used to represent properties, objects, events and relationships between them.
They are combined into deep (multi-layer) overlapping compositional hierarchies. These are artificial neural networks.



Here is a binon structure diagram that shows how binons are combined.
They are binary because they only have links to two lower binons.
The ones in this diagram are entity binons. Two other types of binons are property and control binons.
Property binons are at lower levels of complexity (0 and -1).
Property values from sensors are used to derive the simplest binons at complexity level 1.
Then these are combined and reused in binary hierarchies at higher levels of complexity.
At lower levels they represent more general things like edges, lines or letters and they represent more specific things at higher levels.
So binons are reused and shared in the representation of object and event types.

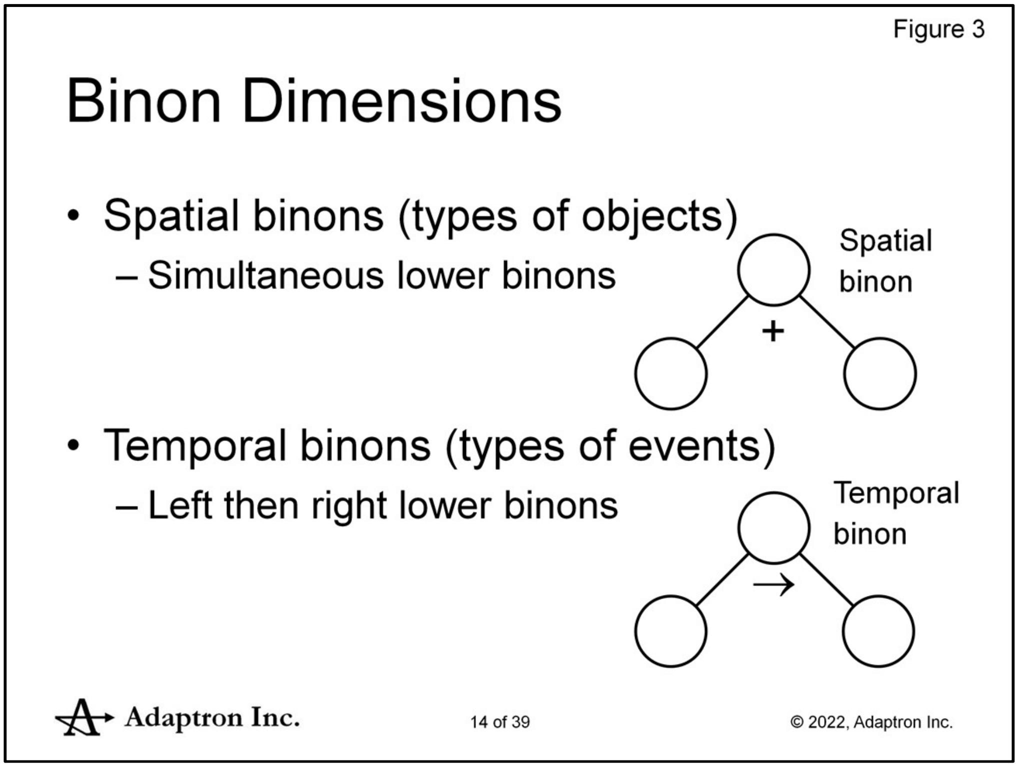


The values in binons are zero or positive integers. Later, I will explain how real numeric values are represented using integers. The values are zero when just the association between the two lower binons is required.

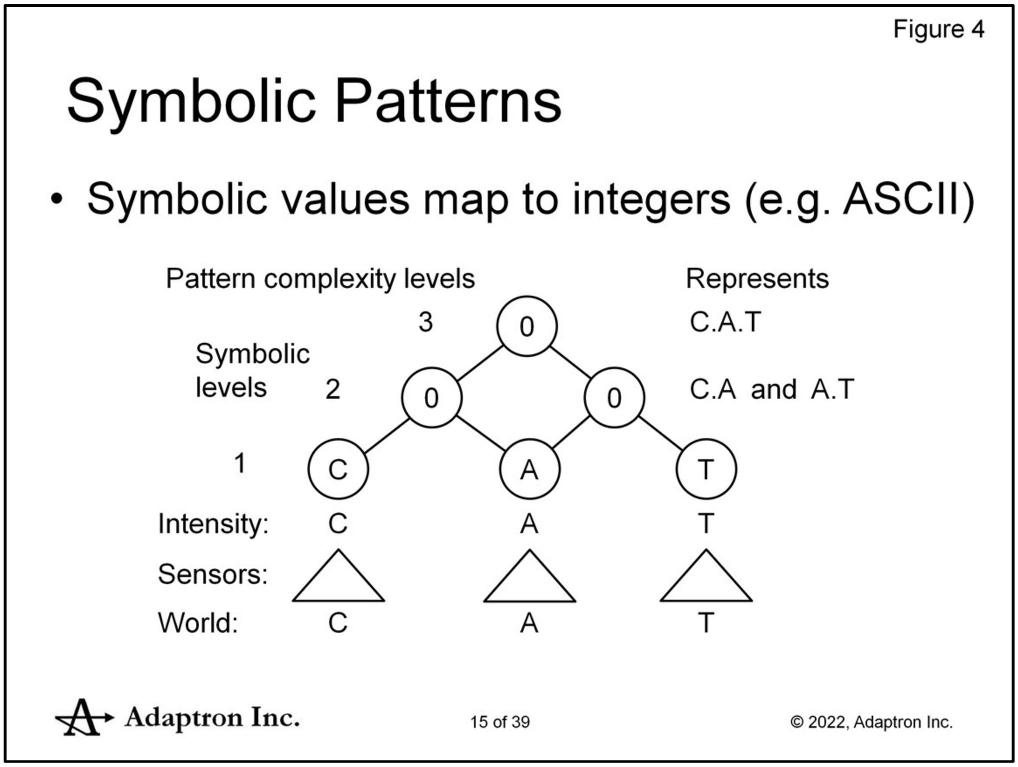
Generalization is top down. Superordinate things are lower down and subordinate things are higher up.

The hierarchies use an overlapping structure so that there is only one unique way of representing any combination and the links are local between layers.

There are no weights on the links and no back propagation is used during learning. If an error occurs then something new has been experienced and a new binon is created to represent it.



Binons are either spatial or temporal so as to represent both object and event types. The lower binons are recognized simultaneously for objects and sequentially for events. The markers of "+" and "→" are used under binons to indicate their dimension.



A symbolic pattern is an easy place to start. Symbolic values such as characters can be represented using integers as in ASCII.
These are perception binons being used for recognizing text.
There are binons for perception, action, mental operation.
There are no dimension markers under the binons because they could be spatial or temporal, depending on the timing of the values.

Perception – Recognizing

- Recognizing
  - Pattern matching with known percepts
  - Bottom-up process
    - Properties and relationships first
    - Then types of objects and events
- Identifying categories of things
- Novel things are encoded and kept in long-term memory

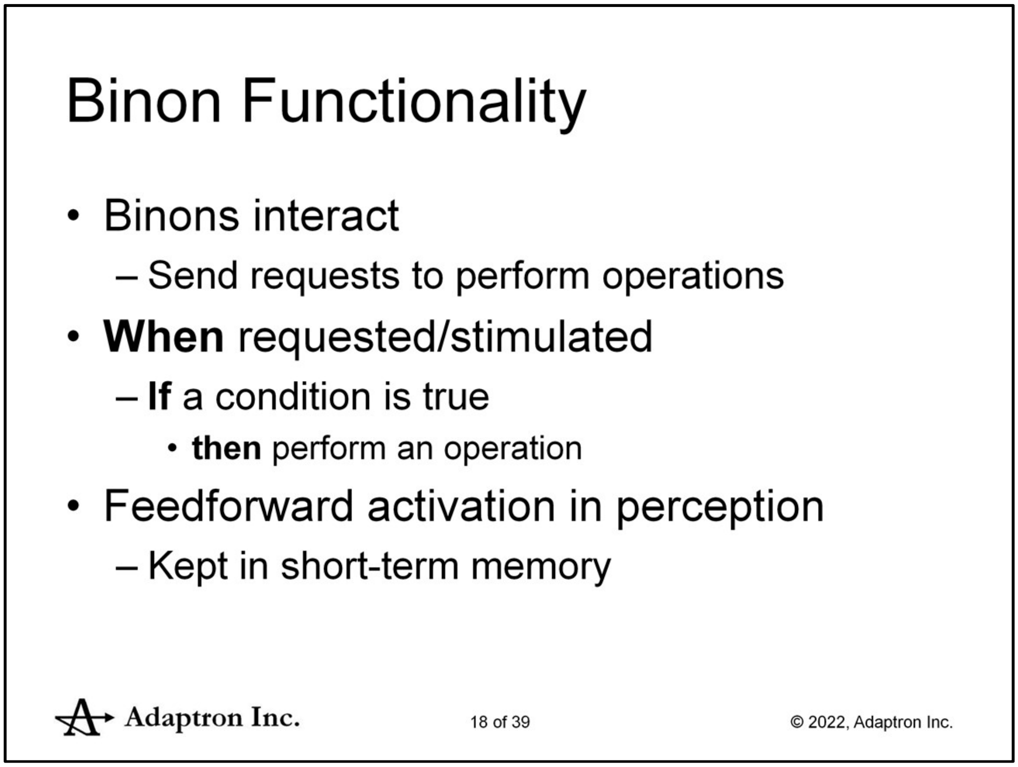
Adaptron Inc.

After sensing comes recognition and encoding.
Recognition is a pattern matching process with known percepts.
It is a bottom-up process that starts with the sensory properties and the relationships derived from them.
These are then combined into symbolic entity binons to represent object and event types.
The result is that they represent categories of things which become more specific at higher levels of complexity.
Novel experiences result in novel binons that are encoded and kept in long-term memory.

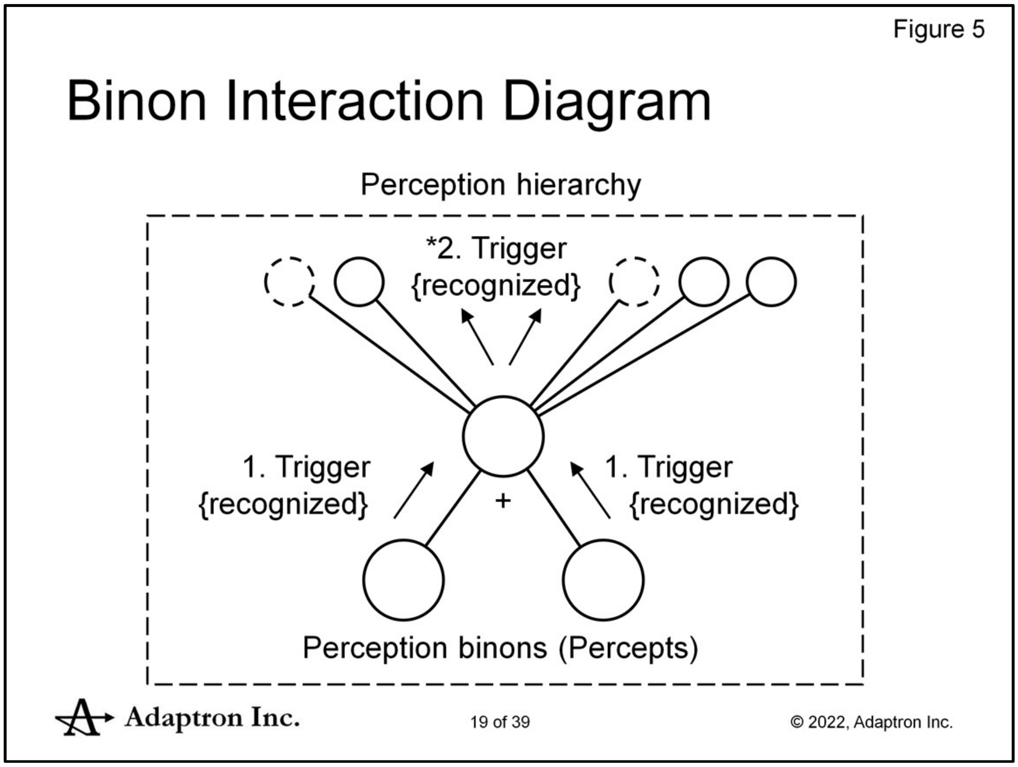
Short-Term Memory (STM)

- Keeps the recognized percepts
  - With their varying sensory properties
  - For counting repetition
- Hierarchical compositional structure
- Two frames
  1. Spatial objects in the current time interval
  2. Event types and their properties

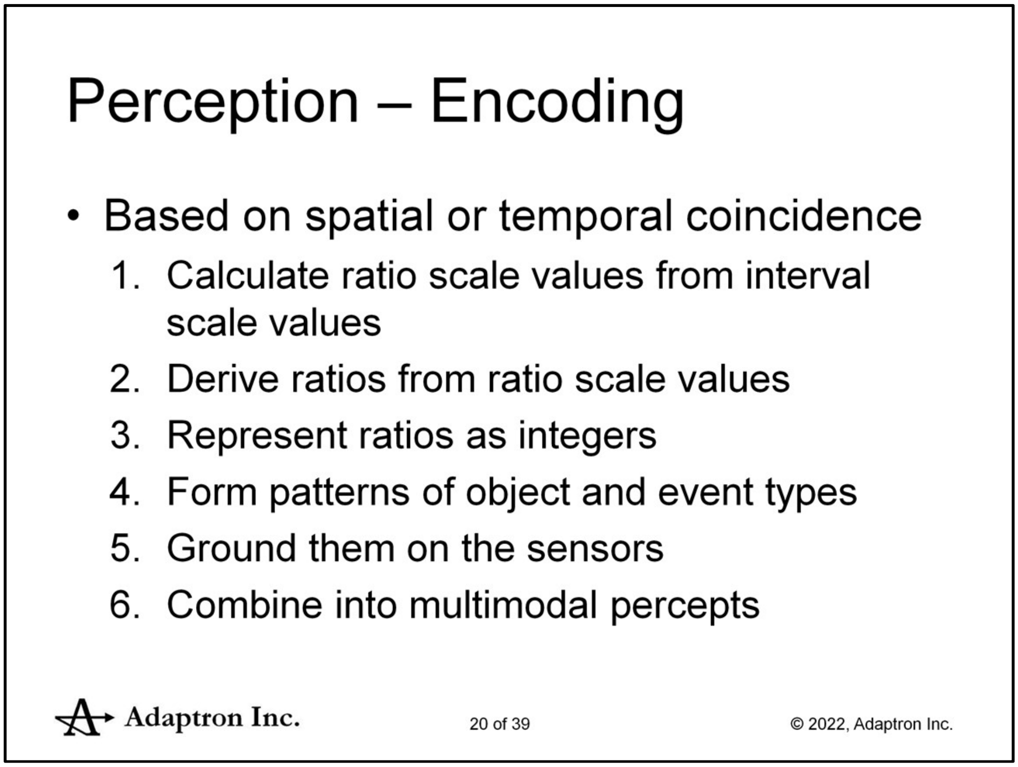
A short-term memory is used to keep track of things as they are being recognized.
The sensory property values are represented in property binons and associated with the entity binons.
STM is used to count objects and events that repeat. This results in the count property.
The structure of STM is the same as the binons in long-term memory except it only contains references to the active ones.
Known percepts for things are recognized first and then novel ones are encoded and added.
There are two frames in STM, one for tracking the binons for types of objects recognized in the current time interval and a second for keeping types of events that occurred due to changes in the first frame.



Binons are functional. They interact with each other like objects in object-oriented software.
They send requests (i.e. messages) to each other to perform operations.
The form of the program is a conditional expression. **When** requested/stimulated **if** a condition is true **then** perform an operation. During perception there is a feedforward wave of these messages that flows up the hierarchies.



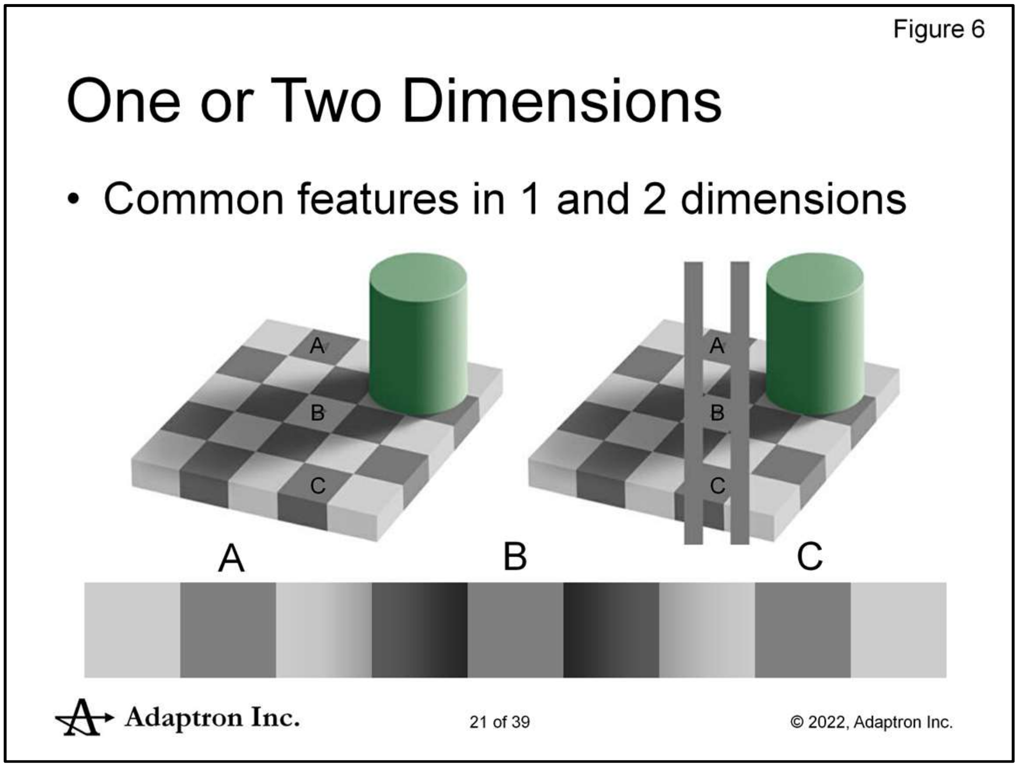
This binon interaction diagram captures the order in which the process occurs. The binon in the middle is waiting for its lower binons to be recognized. When both lower binons are recognized simultaneously they trigger (1) the spatial binon. It then notifies (*2) all of its upper binons. If the binon in the middle was a temporal binon then it would only recognize what it represents when the lower binons are recognized sequentially.
Solid circles are familiar binons and dashed ones are novel binons.



The encoding process for new binons is based on spatial and temporal coincidence detected in the STM.

The steps involved in encoding a percept to represent an object or event type are:

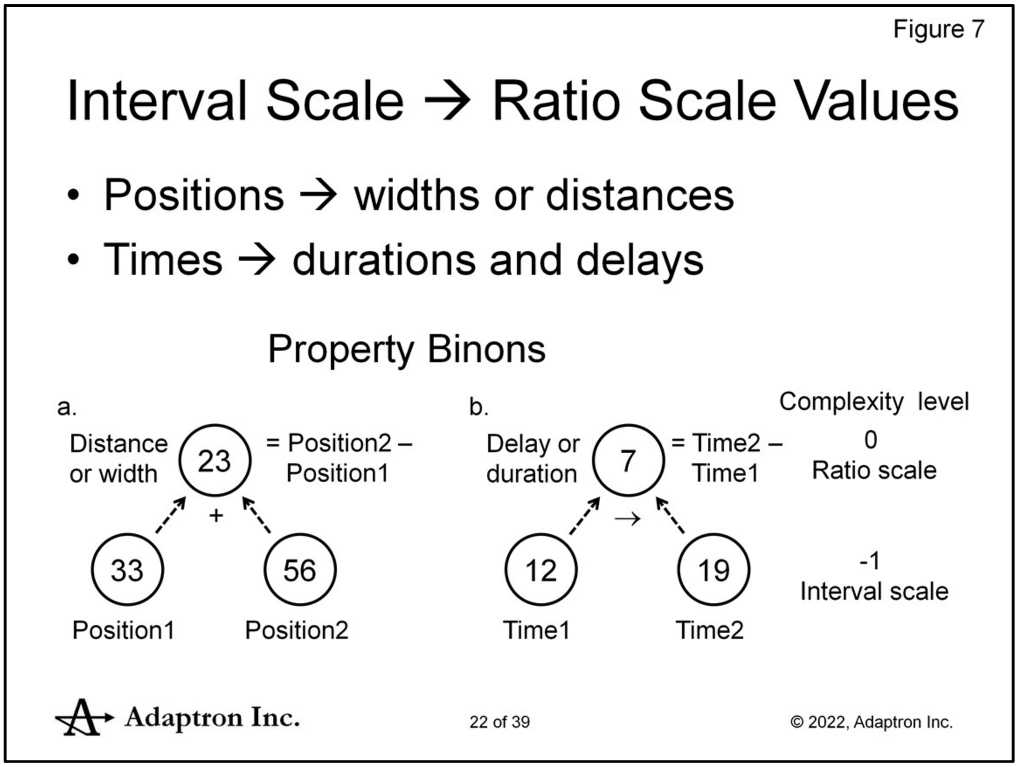
1. Calculate ratio scale values from interval scale values
2. Derive ratios from ratio scale values
3. Represent ratios as integers
4. Form patterns of object and event types
5. Ground them on the sensors
6. Combine into multimodal percepts



Many of the features that occur in two and three dimensions also occur in one dimension. This is the checker shadow illusion. It is an optical illusion published by Edward H. Adelson, Professor of Vision Science at MIT in 1995.

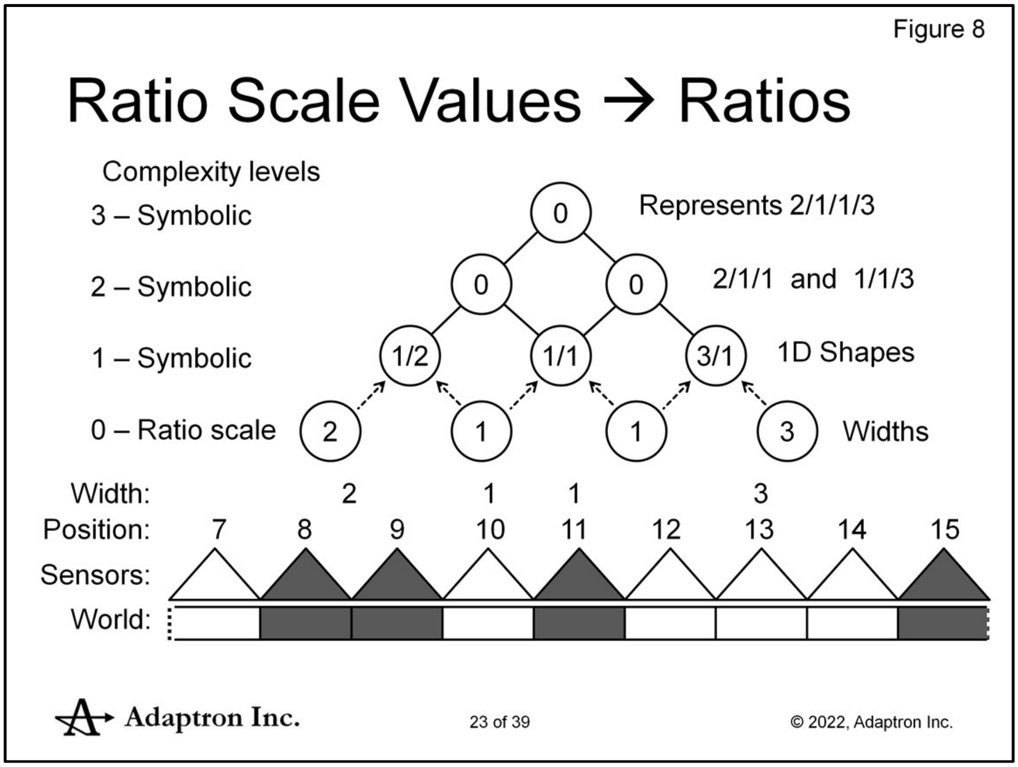
Squares A, B, and C are all the same greyscale intensity value. The gradual changes in intensity are less noticeable than sharp changes at edges.

Therefore, the examples that follow are based on perception using a one dimensional sensory array.

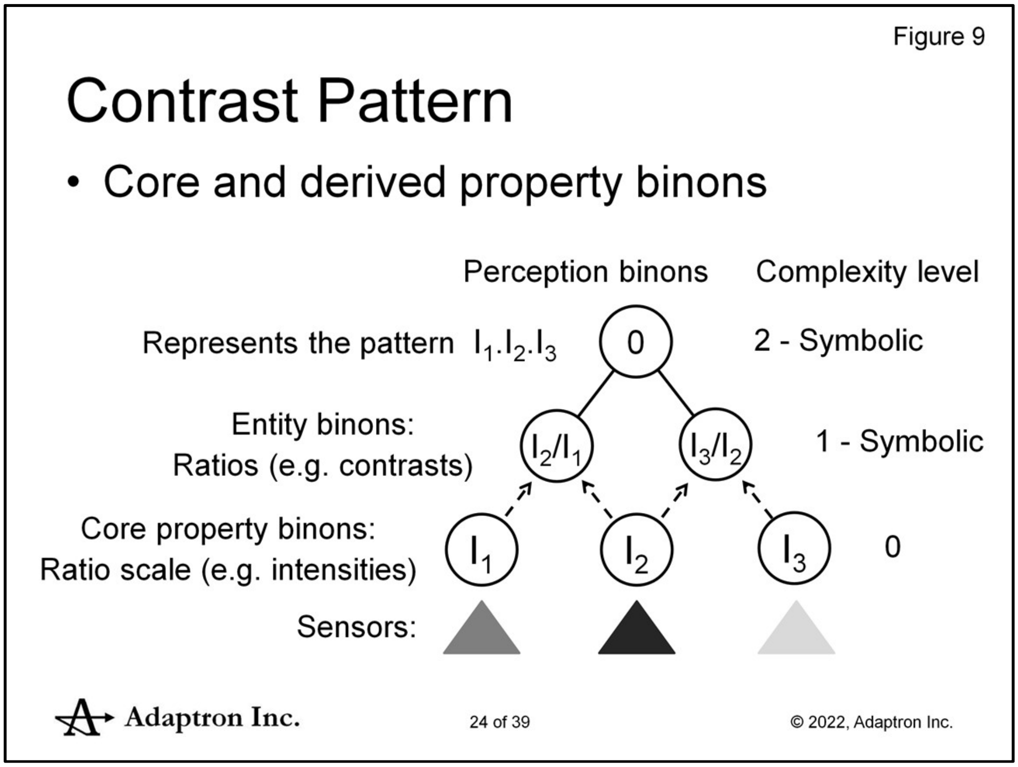


1. Ratio scale value properties are derived from interval scale values by using subtraction. Distance and widths are the differences between positions and durations and delays are the differences between two times. The positions and times are interval scale values in property binons at complexity level -1.

They are used to derive the ratio scale values at complexity level zero. The property binons are not linked together. The dashed arrows indicate they are derived.



2. Derive ratios from ratio scale values
This binon structure diagram shows how the one dimensional shape of 2/1/1/3 is represented.
The ratios are derived from ratio scale properties by using division.
They are calculated by dividing the lower right property by the lower left one.
This is to make it consistent with the derivation of ratio scale values from interval scale values which uses subtraction.
This will become more apparent when we see how ratios are represented as integers.



This binon structure diagram illustrates how the contrast pattern of $I_1.I_2.I_3$ can be represented.
The sensor intensity readings are in the property binons at level zero.
They are used to derive the ratio values in the symbolic entity binons at level 1. And these are combined to form the level 2 entity binon. Its value is zero because only the association of the two level 1 binons needs to be represented.

Repetition [8]

- Objects and events repeat
  - At all levels of complexity
  - Count (spatial and temporal)
    - A core property
    - Ratio scale
    - Numerosity in cognitive science

Adaptron Inc. 25 of 39

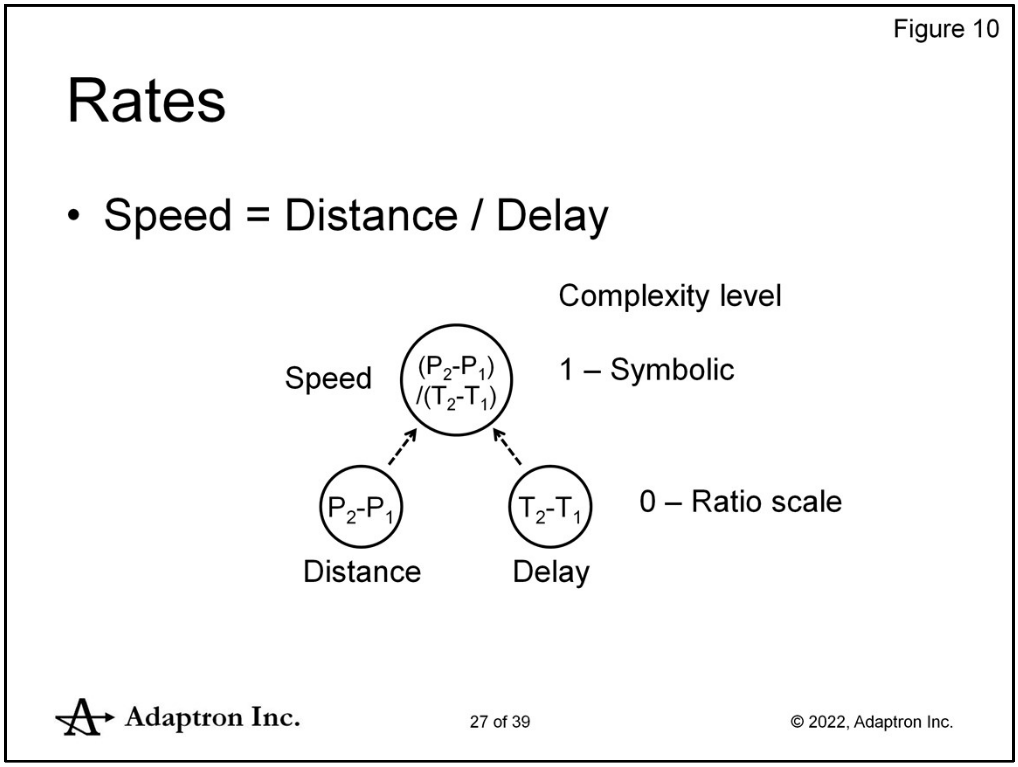
Objects and events repeat and this occurs at all level of complexity.
The STM is used to count them and provide it as an associated core property.
Counts are a core property that can be associated with any object or event that repeats.
Counts are ratio scale values.
This is known as numerosity in cognitive science.

Rates are Ratios [9]

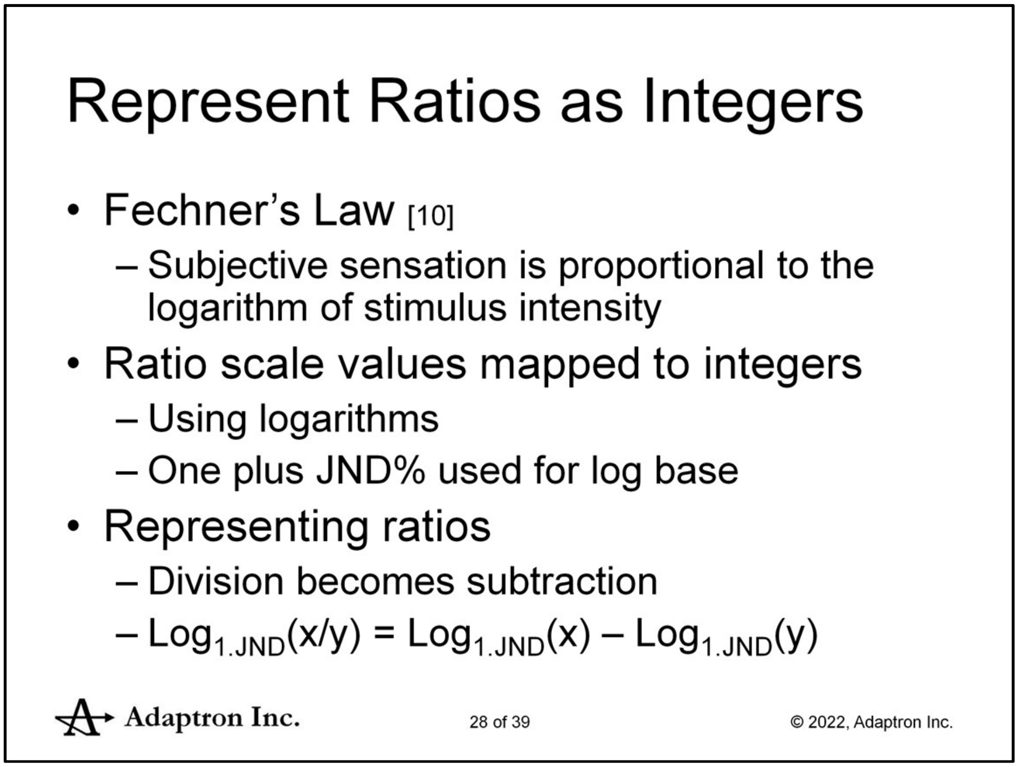
- Spatial properties divided by temporal ones
  - Distance / delay → speed
- Count / duration → temporal frequency
  - E.g. Morse code speed
- Count / width → spatial frequency
  - E.g. knot count in a Persian rug

Adaptron Inc. 26 of 39

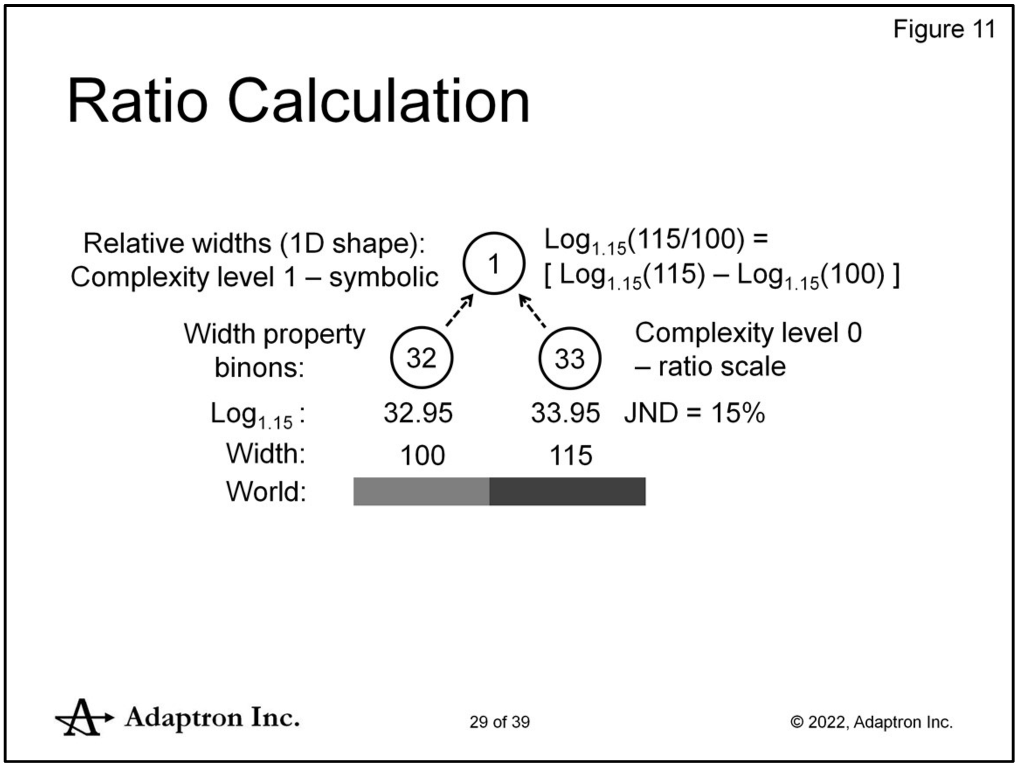
Rates are ratios in which a spatial property is divided by a temporal one.
For example speed is the distance something moved divided by time delay it took.
When a repeat count is divided by a duration you have a temporal frequency such as the speed of Morse code.
When a repeat count is divided by a width you have a spatial frequency such as the knot count in a Persian rug.



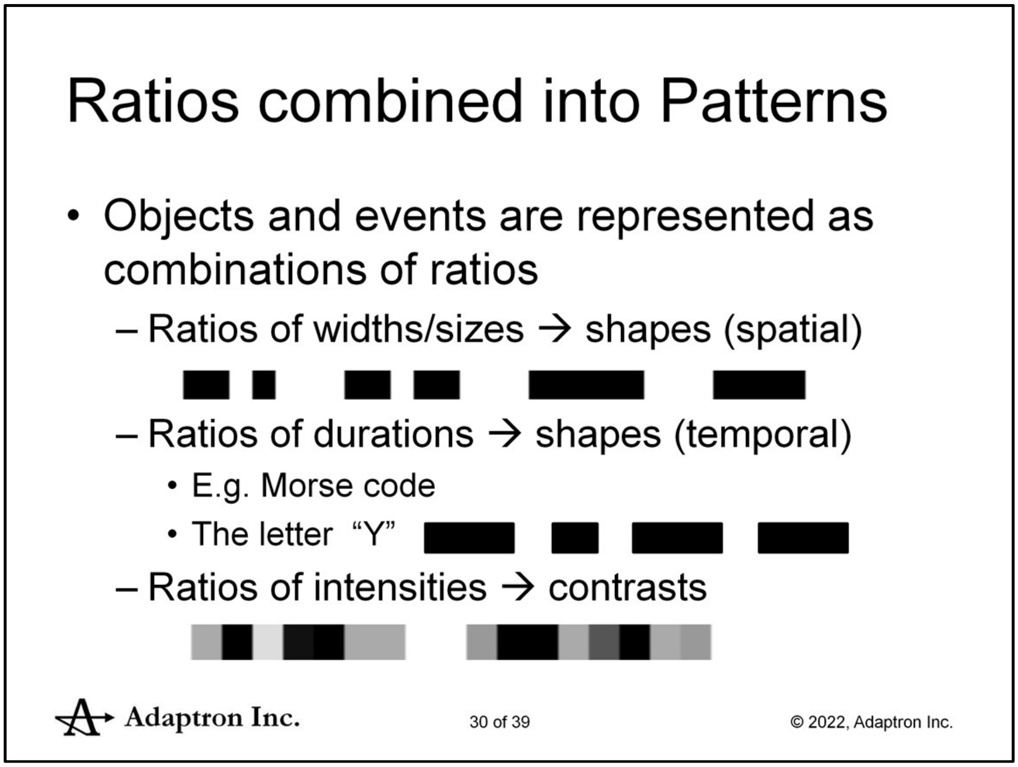
This binon structure diagram illustrates how speed is derived from a distance and a delay which in turn are derived from two positions and two times.



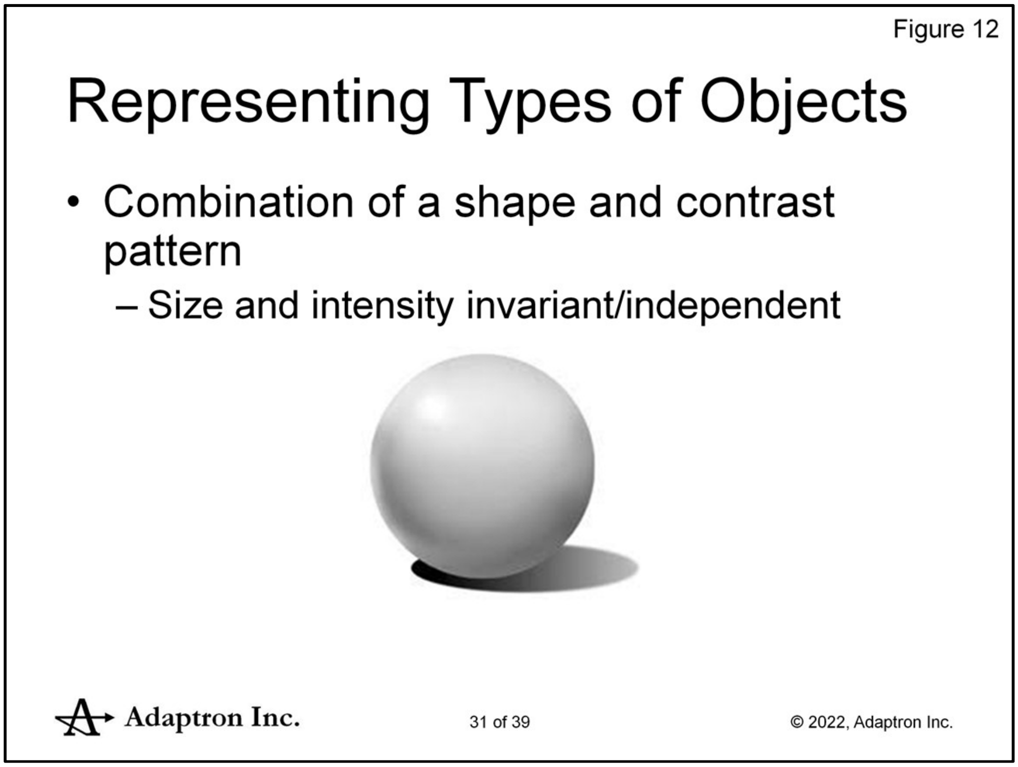
3. Represent ratios as integers.
By using a combination of logarithms and the JND, ratios can be converted into integers
Fechner’s law says that subjective sensation is proportional to the logarithm of stimulus intensity.
By taking the logarithm of a ratio the calculation can be converted from a division into a subtraction.
Log(x/y) = Log(x) – Log(y).
It can then be shown when the two values x and y are not noticeably different and a log base of one plus the JND percentage is used then the integer value of the Log(x/y) will produce the value zero and it will be non zero if they are noticeably different.



Here is an example for two width that are just noticeably different based on a 15% JND.
The integer part of the log of the widths is kept in the level 0 property binons.
The integer part of the difference between the log values represents the ratio 115/100 in the level 1 entity binon.



Ratio are symbolic. They represent a category in which the two source values have a specific relationship to each other.
By combining ratios one can represent shapes, spatial or temporal.
Ratios of intensities can be used to represent contrast patterns which can also be spatial or temporal.



Object types are represented by a combination of their shape and contrast patterns.
They are independent of their size and intensity.
A different contrast pattern would not allow the ball to look spherical.

Representing Types of Events

- Events are patterns of shorter events
  - Piano tune played at different speeds
  - A temporal shape pattern
  - Time scale invariant
- Volume envelope of a musical note
  - Attack, sustain, decay, release
  - A temporal contrast pattern

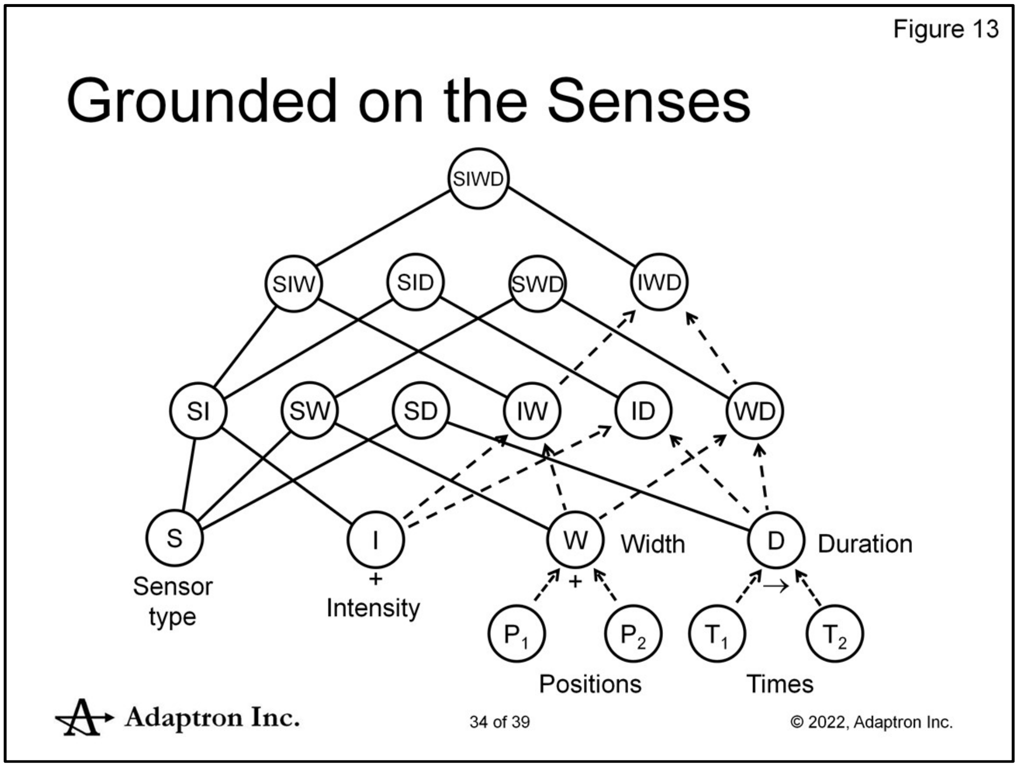
When the notes of a tune are played at different speed you have a temporal shape pattern. The happy birthday tune can be recognized independently of the speed at which it is played. Such patterns of events are time scale invariant.
The volume envelope of a musical note which is composed of its attack, sustain, decay and release allows it to be recognized. It is a temporal contrast pattern.

Types of Objects and Events

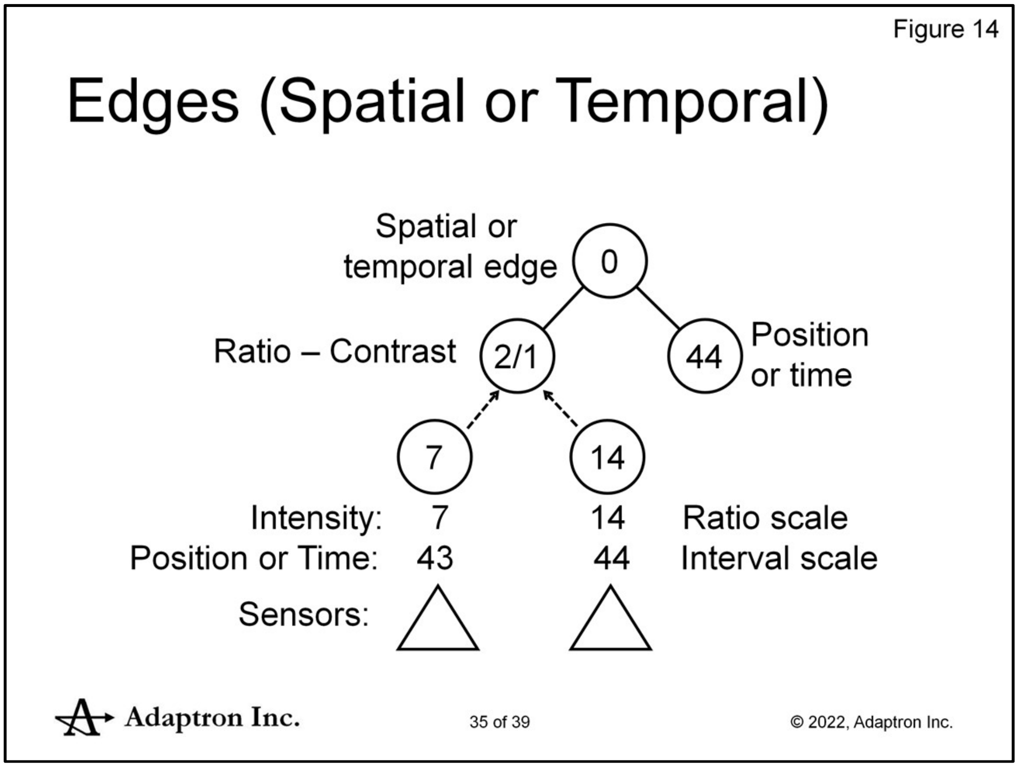
- Properties vary at all levels of complexity
  - Position
  - Size
  - Intensity
  - Time
  - Duration
  - Repetition count

Adaptron Inc. 33 of 39

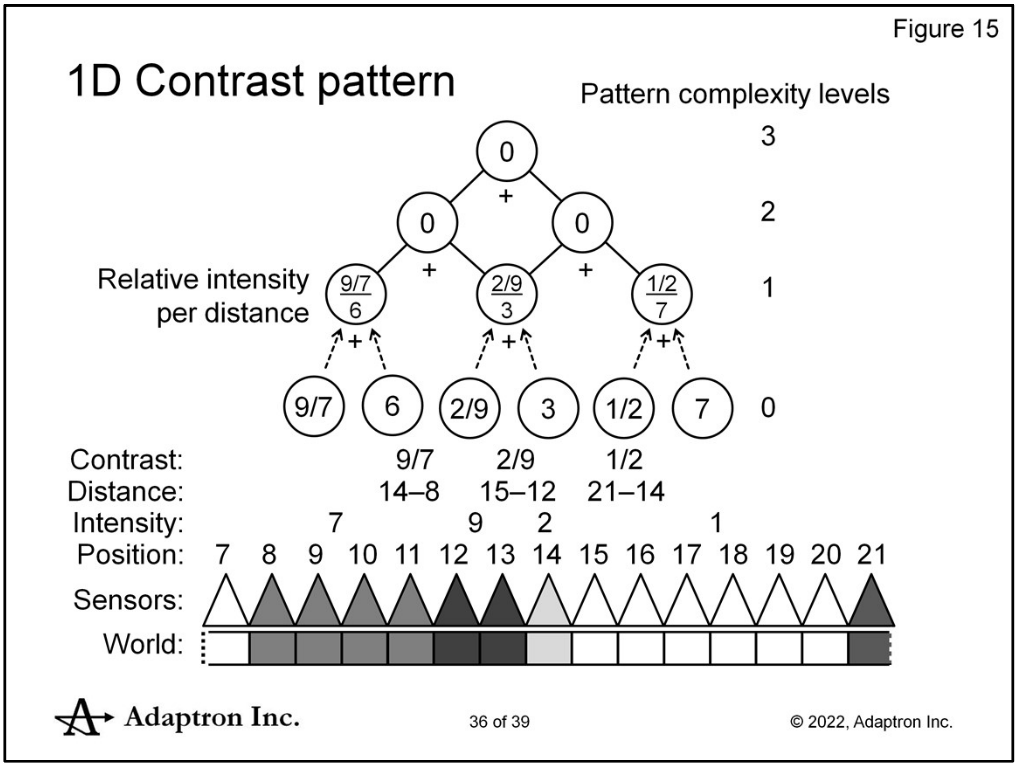
These core and derived properties are variable at all levels of complexity for object and event types.



Object and event types need to be grounded on the sensors because their position, intensity and times can all vary.
They are grounded when they are associated with a sense and sensor type.
Only those properties that change need to be associated with an object type.



An edge is a contrast at a particular position. An edge is the smallest/simplest thing that can be perceived to have moved. This edge is position/time dependent.



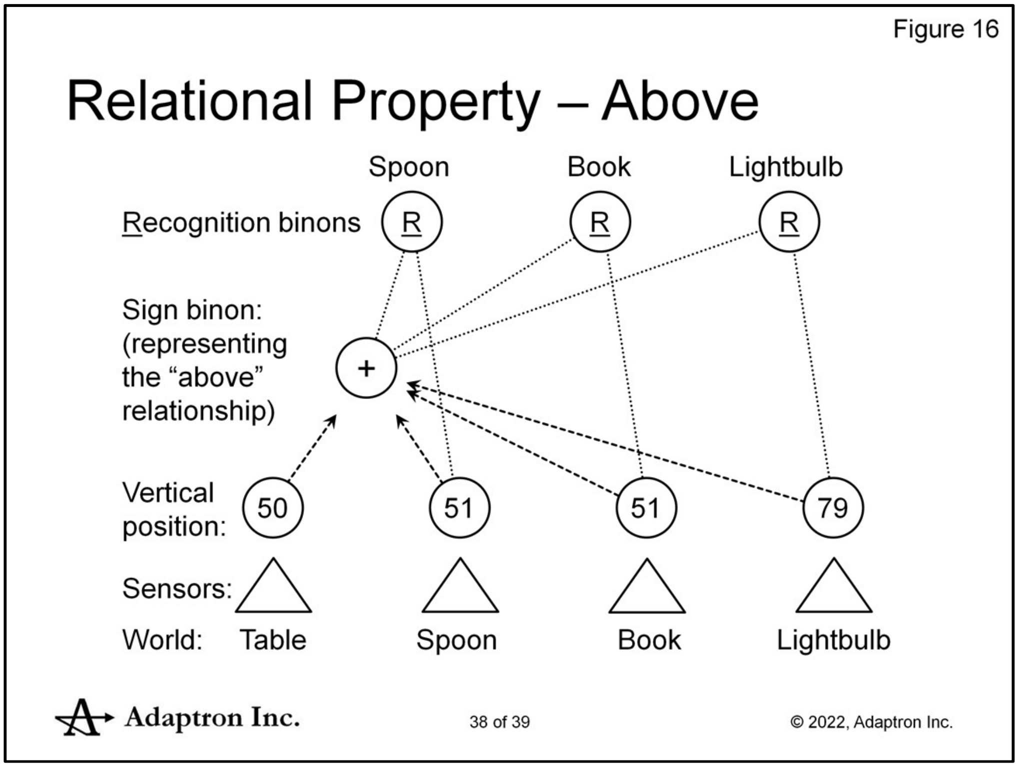
A contrast pattern captures the change of intensity over a distance. But it needs to be position and width independent.

The first contrast between positions 11 and 12 is 9/7 but it spans a distance of 6 = positions 14-8.

Encoding Relational Properties

- Using the sign of
  - Ratio scale values
  - The log of ratios
- Positive, zero or negative
  - Above, on/in, under (position → distance)
  - Before, during, after (time → delay)
  - Bigger, same, smaller (width/size)
  - Louder, same, quieter (sound intensity)
  - Longer, same, shorter (duration)
  - More, equal, less (repeat count)

After calculating the difference between two interval scale values the result is a ratio scale value that maybe positive, zero or negative.
This sign is a symbolic feature that indicates whether something is above on or under for distances and before, during or after for delays
After deriving the log of a ratio the result may be negative or positive if the ratio is below or above one.
This can be used to represent relational features. For example, bigger, same and smaller in size,
louder, same or quieter in sound intensity, longer, same or shorter in duration and the more, equal or less number of things.



This shows how the sign binon is derived from two positions and then associated indirectly as a property of the three objects that are above a table. The dotted lines for the links indicate indirect compositions.

References

[1] Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology*, 53(4), 203-221.

[2] Fidler, S., & Leonardis, A. (2007). Towards scalable representations of object categories: Learning a hierarchy of parts. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007)*.

[3] Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335-346.

[4] Martensen, B. N. (2013). Perceptra: A new approach to pattern classification using a growing network of binary neurons (binons). In R. West & T. Stewart (Eds.), *Proceedings of the 12th International Conference on Cognitive Modeling (ICCM 2013).*

[5] Hammer, B. (2003). Compositionality in neural systems. In M. A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks – 2nd ed.* Cambridge, Massachusetts: MIT Press.

[6] Leslie, A. M., Gallistel, C. R., & Gelman, R. (2019). Where integers come from. In P. Carruthers (Ed.), *The Innate Mind: Foundations and Future.* Oxford: Oxford University Press.

[7] Marcus, G. (2018). Deep learning: A critical appraisal. *Computing Research Repository (CoRR), arXiv:1801.00631v1 [cs.AI].*

[8] Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307-314.

[9] Gallistel, C. R., & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review*, 107(2), 289-344.

[10] Portugal, R. D., & Svaiter, B. F. (2011). Weber-Fechner law and the optimality of the logarithmic scale. *Minds and Machines*, 12(1), 73-81.