BIG DATA AND THE SP THEORY OF INTELLIGENCE

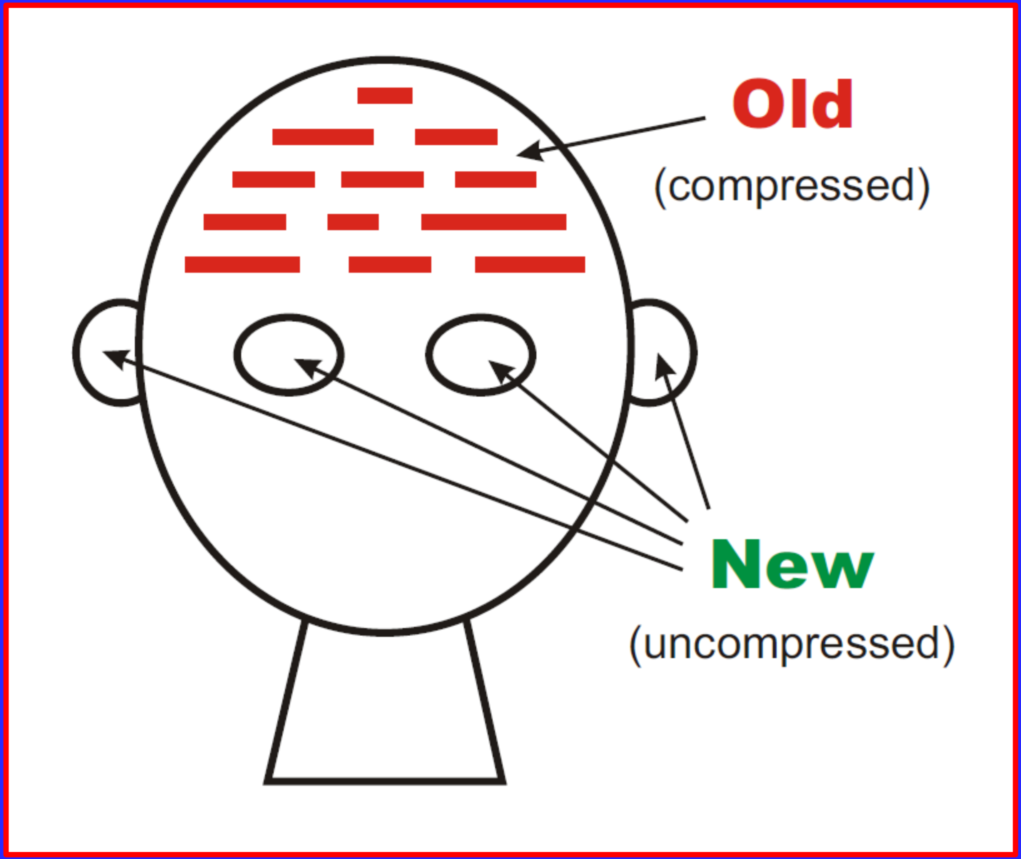
Dr Gerry Wolff

CognitionResearch.org

OVERVIEW

- Outline of the SP theory of intelligence.
- Problems with big data and potential solutions:
  - Volume: big data is … BIG!
  - Efficiency in computation and the use of energy.
  - Unsupervised learning: discovering ‘natural’ structures in data.
  - Transmission of information and the use of energy.
  - Variety: in kinds of data, formats, and modes of processing.
  - Veracity: errors and uncertainties in data.
  - Interpretation of data: pattern recognition, reasoning, …
  - Velocity: analysis of streaming data.
  - Visualisation: representing structures and processes.
- A road map.

OUTLINE OF THE SP THEORY (1)

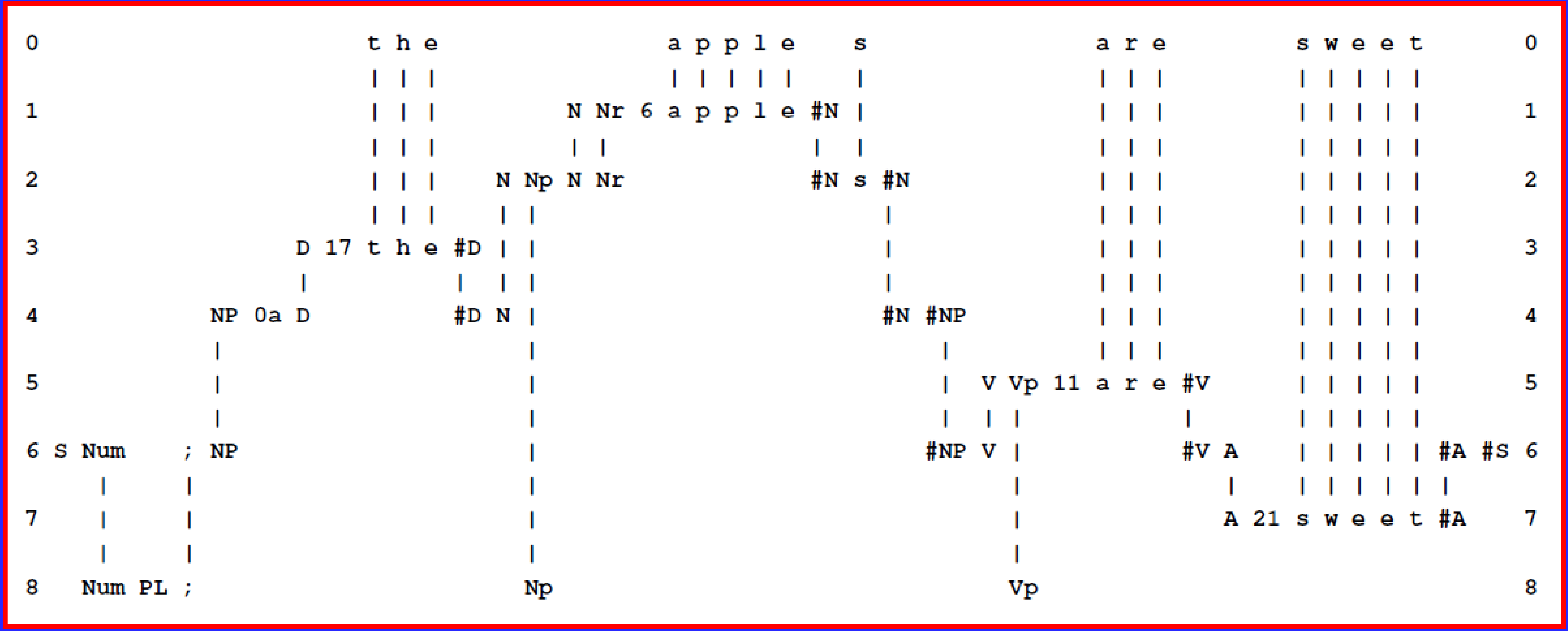


The SP theory is conceived as a brain-like system that receives **New** information and compresses it to create **Old** information .

OUTLINE (2)

- The SP theory is designed to simplify and integrate concepts across artificial intelligence, mainstream computing, and human perception and cognition.
- The SP theory is the product of an extensive programme of development and testing via the SP computer model. This model also serves for demonstrating what can be done with the system.
- **All** kinds of knowledge are represented with arrays of atomic symbols in one or two dimensions. These are called “patterns”.
- **All** kinds of processing are done by compressing information:
  - Via the matching and unification of patterns.
  - More specifically, via the building of multiple alignments (next).

AN SP MULTIPLE ALIGNMENT



```
0                       t h e                a p p l e    s                a r e         s w e e t       0
                        | | |                | | | | |    |                | | |         | | | | |
1                       | | |         N Nr 6 a p p l e #N |                | | |         | | | | |       1
                        | | |         | |              |  |                | | |         | | | | |
2                       | | |    N Np N Nr             #N s #N             | | |         | | | | |       2
                        | | |    | |                        |              | | |         | | | | |
3                  D 17 t h e #D | |                        |              | | |         | | | | |       3
                   |          |  | |                        |              | | |         | | | | |
4            NP 0a D          #D N |                        #N #NP         | | |         | | | | |       4
             |                     |                            |          | | |         | | | | |
5            |                     |                            |  V Vp 11 a r e #V      | | | | |       5
             |                     |                            |  | |           |       | | | | |
6 S Num    ; NP                    |                           #NP V |           #V A    | | | | | #A #S 6
     |     |                       |                                 |              |    | | | | | |
7    |     |                       |                                 |              A 21 s w e e t #A    7
     |     |                       |                                 |
8   Num PL ;                       Np                                Vp                                  8
```

CognitionResearch.org

OUTLINE (3)

- Prediction and probabilities: the SP system is fundamentally probabilistic because of the close relation between information compression and probability.
- Versatility: The SP system has strengths in several areas.
- Simplification and seamless integration of structures and functions because:
  - All kinds of knowledge are represented with patterns.
  - All kinds of processing are done via the creation of multiple alignments.
- SP-neural: The SP system may be realised with neurons.

VOLUME

- Problem: “Very-large-scale data sets introduce many data management challenges.” (National Research Council, *Frontiers in Massive Data Analysis*, National Academies Press, 2013, p. 41).
- Solutions: By compressing big data, the SP system yields:
  - Direct benefits in storage, management and transmission.
  - Indirect benefits via efficiency in computation and the use of energy, via unsupervised learning, via additional economies in transmission and the use of energy, via assistance in the management of errors and uncertainties in data, and via processes of interpretation.

EFFICIENCY IN COMPUTATION AND THE USE OF ENERGY (1)

- Moore’s Law: “... we’re reaching the limits of our ability to make [gains in the capabilities of CPUs] at a time when we need even more computing power to deal with complexity and big data. And that's putting unbearable demands on today’s computing technologies—mainly because today’s computers require so much energy to perform their work.” (John E Kelly III & Steve Hamm, *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. New York: Columbia University Press, 2013, p. 9).

- Energy: “The human brain is a marvel. A mere 20 watts of energy are required to power the 22 billion neurons in a brain that’s roughly the size of a grapefruit. To field a conventional computer with comparable cognitive capacity would require gigawatts of electricity and a machine the size of a football field. So, clearly, something has to change fundamentally in computing for sensing machines to help us make use of the millions of hours of video, billions of photographs, and countless sensory signals that surround us. ... Unless we can make computers many orders of magnitude more energy efficient, we’re not going to be able to use them extensively as our intelligent assistants.” (Kelly & Hamm, pp. 75 & 88).

EFFICIENCY (2)

- In the SP theory, a process of searching for matching patterns is central in **all** kinds of ‘processing’ or ‘computing’.
- This means that anything that increases the efficiency of searching will increase computational efficiency and, probably, cut the use of energy:
  - Reducing the volume of big data.
  - Exploiting *****probabilities*****.
  - Cutting out some searching.
- There may also be savings:
  - Via a synergy with data-centric computing.
  - Via efficiency in transmission of information (later).

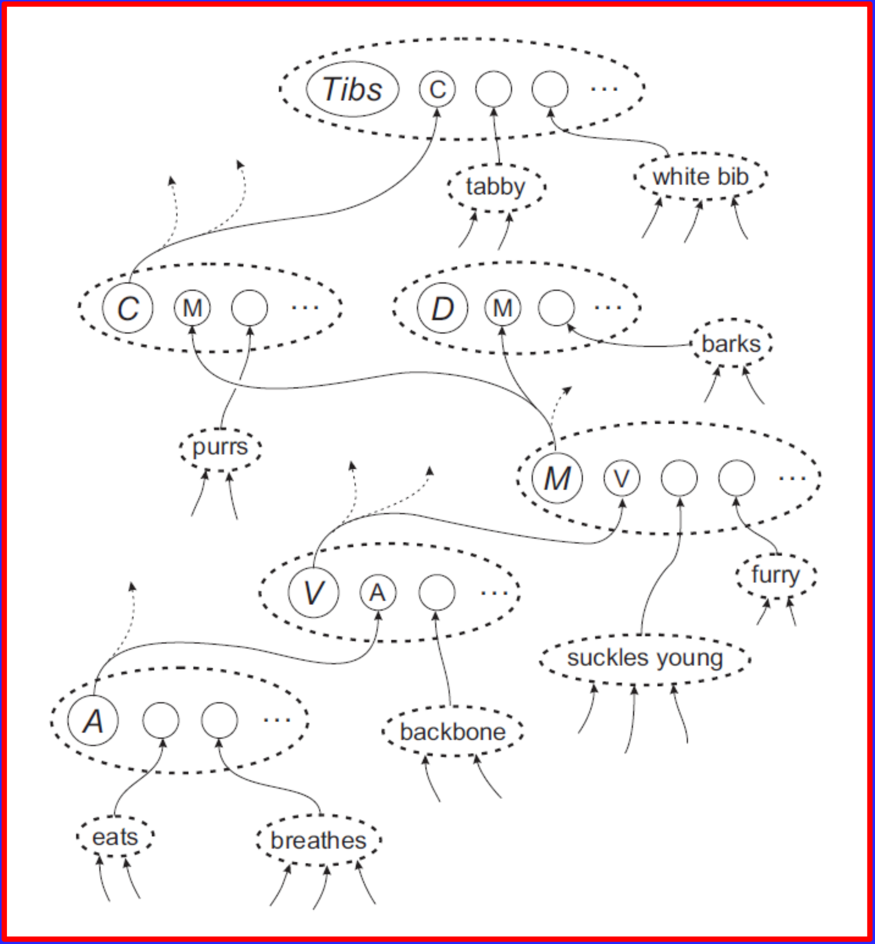
EFFICIENCY VIA REDUCTIONS IN VOLUME

- Information compression is central in how the SP system works:
  - Reducing the size of big data.
  - Reducing the size of search terms.
- Both these things can increase the efficiency of searching, meaning gains in computational efficiency and cuts in the use of energy.

EFFICIENCY VIA PROBABILITIES

- “If we want to find some strawberry jam, our search is more likely to be successful in a supermarket than it would be in a hardware shop or a car-sales showroom.”
- Statistical knowledge flows directly from:
  - Information compression in the SP system and
  - The intimate connection between information compression and concepts of prediction and probability.
- There is great potential to cut out unnecessary searching, with consequent gains in efficiency.
- Potential for savings at all levels and in all parts of the system and on many fronts in its stored knowledge.

EFFICIENCY VIA A SYNERGY WITH DATA-CENTRIC COMPUTING AND CUTTING OUT SOME SEARCHNG



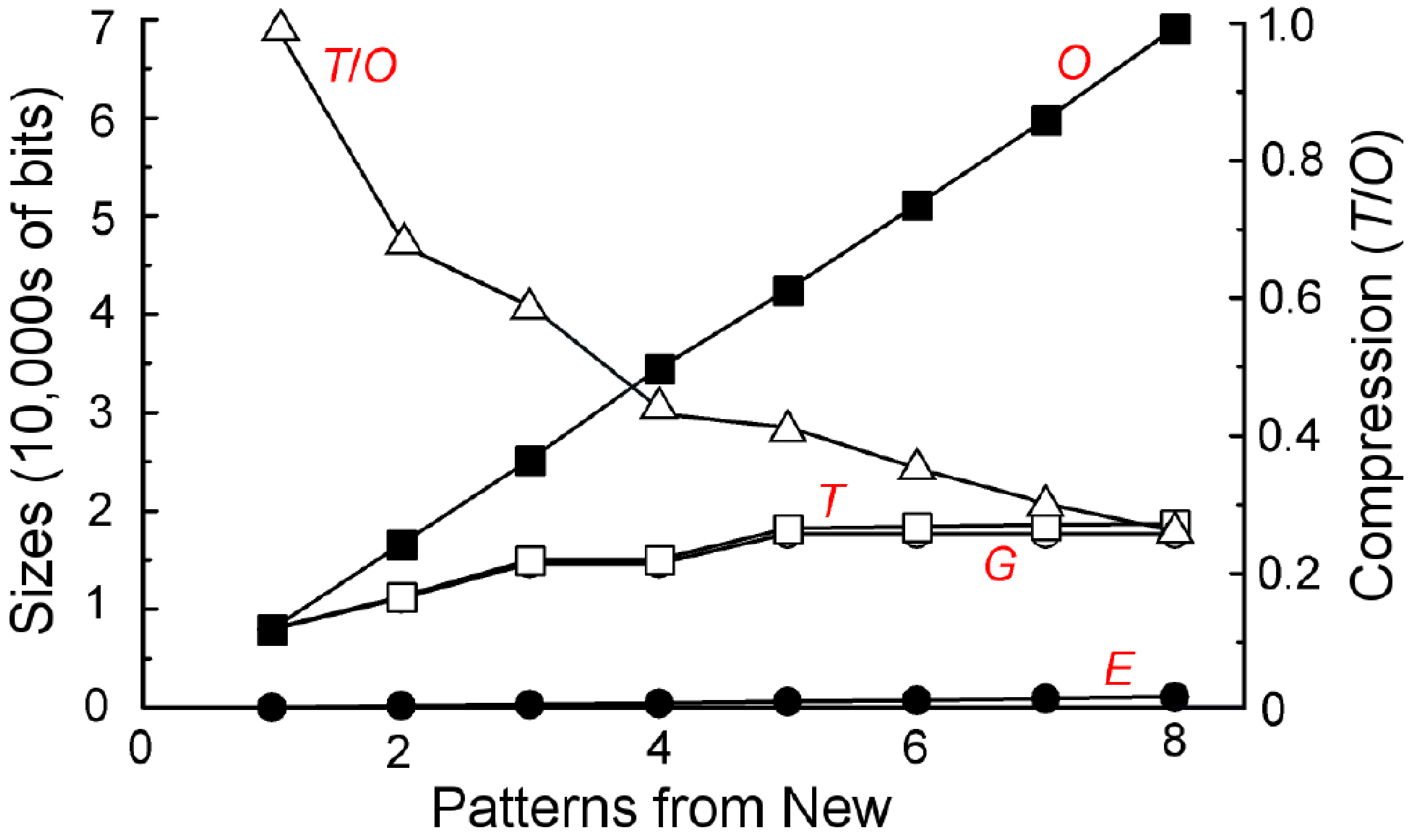
- In *SP-neural*, SP patterns may be realised as neuronal *pattern assemblies*.
- There would be close integration of data and processing, as in data-centric computing (Kelly & Hamm).
- Direct connections may cut out some searching.

CognitionResearch.org

UNSUPERVISED LEARNING

- “While traditional computers must be programmed by humans to perform specific tasks, cognitive systems will learn from their interactions with data and humans and be able to, in a sense, program themselves to perform new tasks.” (Kelly & Hamm, p. 7).
- The SP programme of research grew out of earlier research developing computer models of language learning.
- But because of the new goals of simplification and integration, a radically new structure has been needed. This is the multiple alignment framework.
- Information compression, or “minimum length encoding” remains the key.
- The SP computer model has already demonstrated an ability to discover generative grammars, including segmental structures, classes of structure, and abstract patterns.
- There is potential to learn other kinds of structure such as class hierarchies, part-whole hierarchies, and their integration.
- From a body of information, I, the products of learning are: a grammar (G) and an encoding (E) of I in terms of G (next).

THE PRODUCTS OF LEARNING



O = original (New) data (cumulative)

E = encoded data (cumulative)

G = grammar (cumulative)

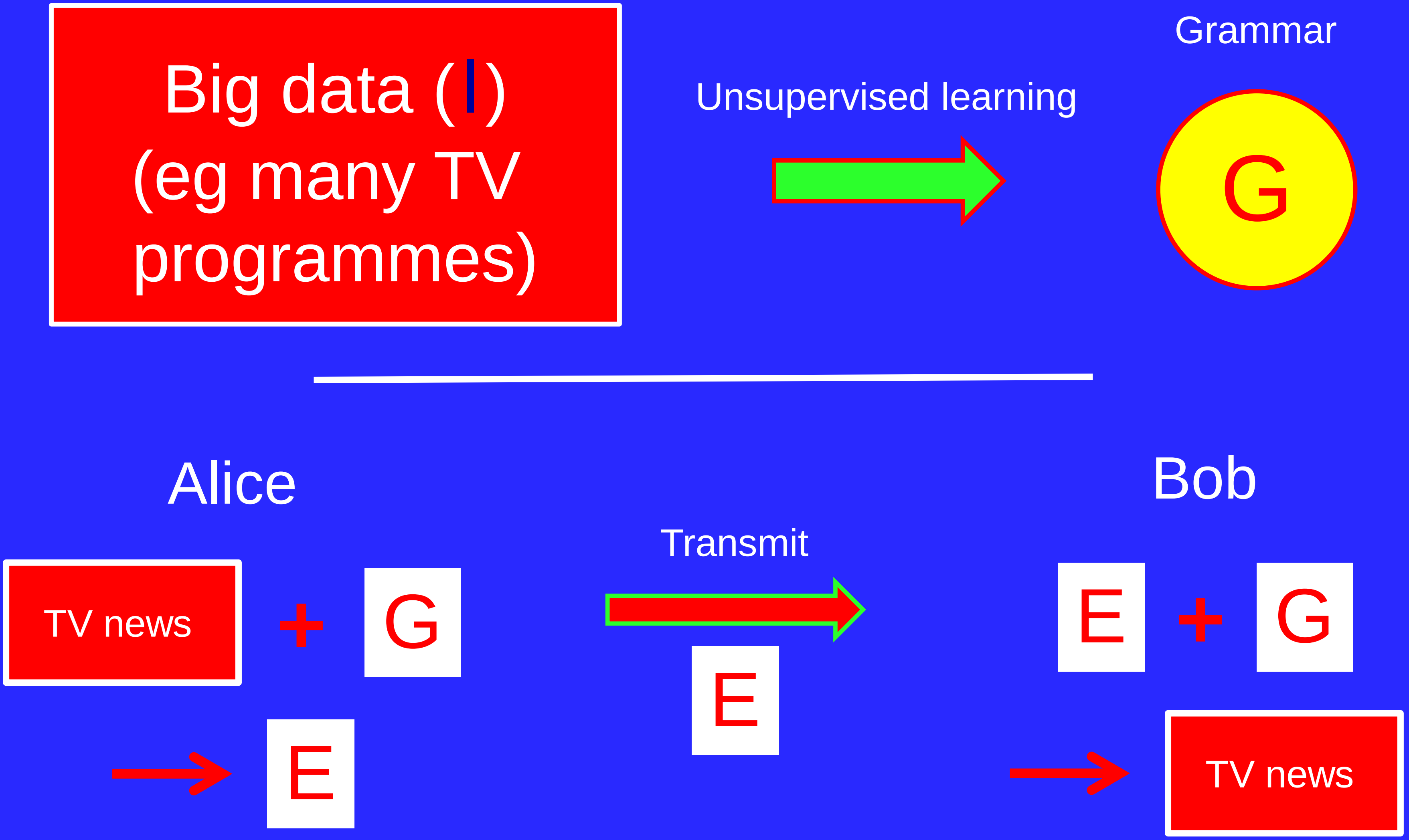
T = *G* + *E*

CognitionResearch.org

TRANSMISSION OF INFORMATION

- “To control costs, designers of the [DOME] computing system have to figure out how to minimize the amount of energy used for processing data. At the same time, since so much of the energy in computing is required to move data around, they have to discover ways to move the data as little as possible.” (Kelly & Hamm, p. 65).
- The SP system can increase the efficiency of transmission:
  - By making big data smaller (“Volume”).
  - By separating grammar (**G**) from encoding (**E**), as in some dictionary techniques and analysis/synthesis schemes (next slide).
- Efficiency in transmission can mean cuts in the use of energy.

SEPARATING GRAMMAR (G) FROM ENCODING (E)



Big data (I) (eg many TV programmes)

Unsupervised learning

Grammar

G

Alice

TV news + G

→ E

Transmit

E

Bob

E + G

→ TV news

POTENTIAL ADVANTAGES OF SP SYSTEM IN TRANSMISSION OF INFORMATION

- The relative simplicity of a focus on the matching and unification of patterns.
- The system aims to discover structures that are, quotes, “natural”. This brain-inspired “DONSVIC” principle can mean relatively high levels of information compression.
- There is potential for G to include structures not recognised by most compression algorithms, such as:
  - Generic 3D models of objects and scenes.
  - Generic sequential redundancies across sequences of frames.

VARIETY (1)

- "The manipulation and integration of heterogeneous data from different sources into a meaningful common representation is a major challenge." National Research Council, *Frontiers in Massive Data Analysis*, National Academies Press, 2013, p 76.
- "Over the past decade or so, computer scientists and mathematicians have become quite proficient at handling specific types of data by using specialized tools that do one thing very well. ... But that approach doesn't work for complex operational challenges such as managing cities, global supply chains, or power grids, where many interdependencies exist and many different kinds of data have to be taken into consideration." (Kelly & Hamm, p. 48).

VARIETY (2)

- Diverse kinds of data: the world’s many languages, spoken or written; static and moving images; music as sound and music in its written form; numbers and mathematical notations; tables; charts; graphs; networks; trees; grammars; computer programs; and more.
- There are often several different computer formats for each kind of data. With images, for example: JPEG, TIFF, WMF, BMP, GIF, EPS, PDF, PNG, PBM, and more.
- Adding to the complexity is that each kind of data and each format normally requires its own special mode of processing.
- THIS IS A MESS! It needs cleaning up.
- Although some kinds of diversity are useful, there is a case for developing a *universal framework for the representation and processing of diverse kinds of knowledge* (UFK).

VARIETY (3)

- Potential benefits of a UFK in: ● Learning structure in data (next slide); ● Interpretation of data; ● Data fusion; ● Understanding and translation of natural languages; ● The semantic web and internet of things; ● Long-term preservation of data; ● Seamless integration in the representation and processing of diverse kinds of knowledge.
- The human brain seems to function as a UFK:
    - Everything is done with neurons;
    - One part of the brain can take over the role of another part;
    - Most concepts are an amalgam of diverse kinds of knowledge (which implies some uniformity in the representation and processing of diverse kinds of knowledge).
- The SP system is a good candidate for the role of UFK because of its versatility in the representation and processing of diverse kinds of knowledge.

HOW VARIETY HINDERS LEARNING

- Discovering the association between lightning and thunder is likely to be difficult when:
  - Lightning appears in big data as a static image in one of several formats; or in a moving image in one of several formats; or it is described, in spoken or written form, as any one of such things as “firebolt”, “fulmination”, “la foudre”, “der Blitz”, “lluched”, “a big flash in the sky”, or indeed “lightning”.
  - Thunder is represented in one of several different audio formats; or it is described, in spoken or written form, as “thunder”, “gök gürültüsü”, “le tonnerre”, “a great rumble”, and so on.
- If learning and discovery processes are going to work effectively, we need to get behind these surface forms and focus on the underlying meanings. This can be done using a UFK.

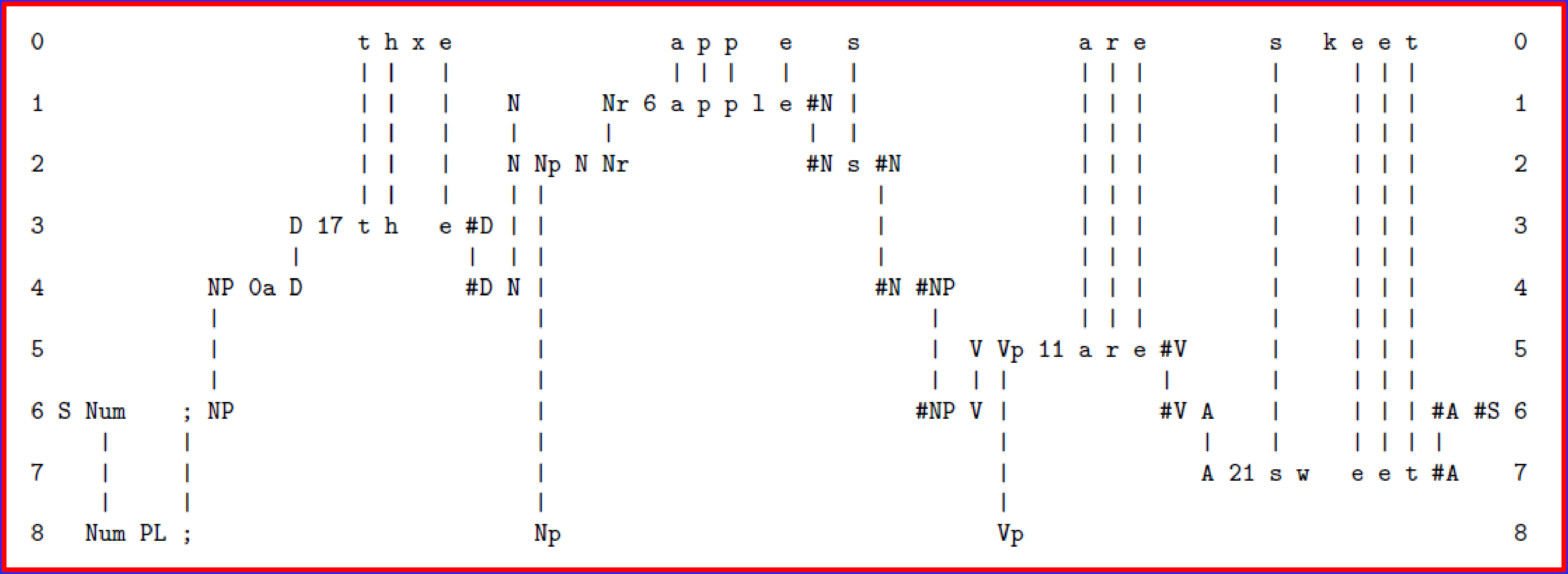
VERACITY (1)

- “In building a statistical model from any data source, one must often deal with the fact that data are imperfect. Real-world data are corrupted with noise. … Measurement processes are inherently noisy, data can be recorded with error, and parts of the data may be missing.” (NRC, p. 99).

- “Organizations face huge challenges as they attempt to get their arms around the complex interactions between natural and human-made systems. The enemy is uncertainty. In the past, since computing systems didn’t handle uncertainty well, the tendency was to pretend that it didn’t exist. Today, it is clear that that approach won’t work anymore. So rather than trying to eliminate uncertainty, people have to embrace it.” [Kelly & Hamm, pp. 50-51].

VERACITY (2)

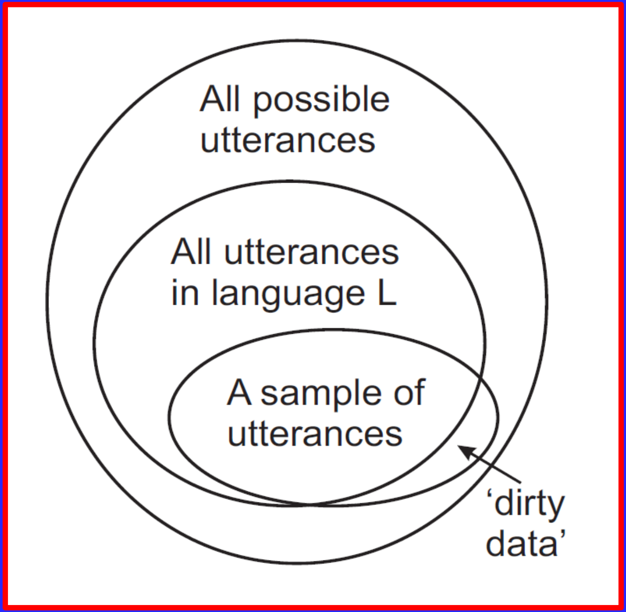
In tasks such as parsing or pattern recognition, the SP system is robust in the face of errors of omission, addition, or substitution.



```
0                       t h x e                a p p   e    s                a r e         s   k e e t       0
                        | |   |                | | |   |    |                | | |         |     | | |
1                       | |   |    N      Nr 6 a p p l e #N |                | | |         |     | | |       1
                        | |   |    |      |              |  |                | | |         |     | | |
2                       | |   |    N Np N Nr             #N s #N             | | |         |     | | |       2
                        | |   |    | |                        |              | | |         |     | | |
3                  D 17 t h   e #D | |                        |              | | |         |     | | |       3
                   |            |  | |                        |              | | |         |     | | |
4            NP Oa D            #D N |                        #N #NP         | | |         |     | | |       4
             |                       |                           |           | | |         |     | | |
5            |                       |                           |   V Vp 11 a r e #V      |     | | |       5
             |                       |                           |   | |           |       |     | | |
6 S Num    ; NP                      |                           #NP V |           #V A    |     | | | #A #S 6
    |      |                         |                                 |              |    |     | | | |
7   |      |                         |                                 |              A 21 s w   e e t #A    7
    |      |                         |                                 |
8   Num PL ;                         Np                                Vp                                    8
```

VERACITY (3)

- When we learn a first language (L):
  - We learn from a finite sample.
  - We generalise (to L) without over-generalising.
  - We learn ‘correct’ knowledge despite ‘dirty data’.
- For any body of data, I, principles of minimum-length encoding provide the key:



  - Aim to minimise the overall size of G and E.
  - G is a distillation or ‘essence’ of I, that excludes most ‘errors’ and generalises beyond I.
  - E + G is a lossless compression of I including typos etc but without generalisations.
  - Systematic distortions remain a problem.

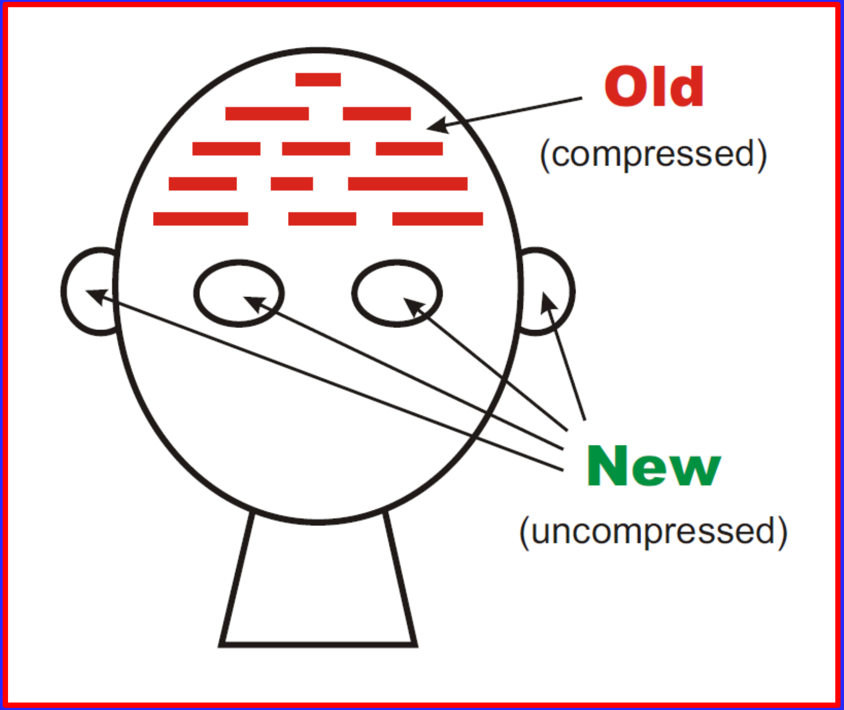
CognitionResearch.org

INTERPRETATION OF DATA

- In the SP system, interpretation of a body of data, I, means processing it in conjunction with a pre-established grammar, G, to create multiple alignments (and E).
- Depending on the nature of I and G, interpretation may be seen as:
  - Pattern recognition;
  - Information retrieval;
  - Parsing or production of natural language;
  - Translation from one representation to another;
  - Scanning big data for pre-defined patterns;
  - One of several different kinds of reasoning;
  - Planning;
  - Problem solving.
- The SP system has strengths in all these areas.

VELOCITY

- In the context of big data, “velocity” means the analysis of streaming data as it is received.
- “This is the way humans process information.” (Kelly & Hamm, p. 50).
- This style of analysis is at the heart of how the SP system has been designed (right).

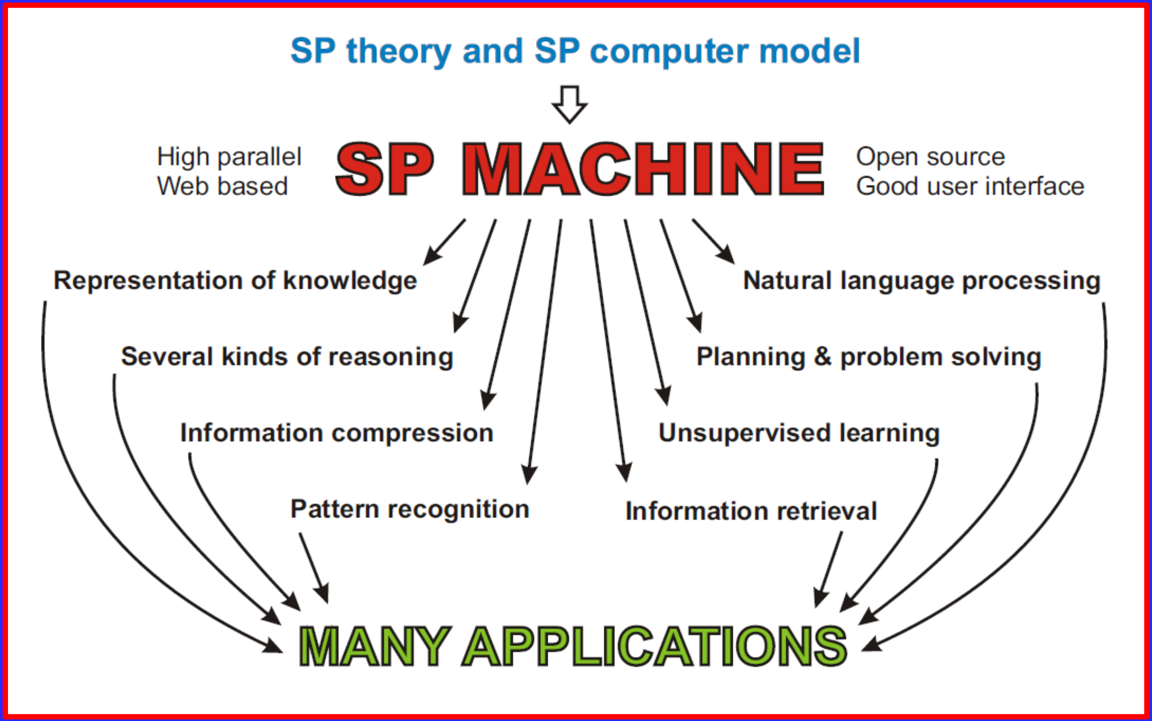


VISUALISATION

- “... methods for visualization and exploration of complex and vast data constitute a crucial component of an analytics infrastructure.” (NRC, p. 133).
- “[An area] that requires attention is the integration of visualization with statistical methods and other analytic techniques in order to support discovery and analysis." [NRC, p. 142].
- The SP system is well suited to visualisation for these reasons:
  - Transparency in the representation of knowledge.
  - Transparency in processing.
  - The system is designed to discover ‘natural’ structures in data.
  - In accordance with what the NRC says, there is clear potential to integrate visualisation with the statistical techniques that lie at the heart of how the SP system works.

A ROAD MAP

- Develop a high-parallel, open-source version of the SP machine.
- This facility would be a means for researchers everywhere to explore what can be done with the system and to create new versions of it.



SP theory and SP computer model

High parallel
Web based

SP MACHINE

Open source
Good user interface

Representation of knowledge

Natural language processing

Several kinds of reasoning

Planning & problem solving

Information compression

Unsupervised learning

Pattern recognition

Information retrieval

MANY APPLICATIONS

CognitionResearch.org

FURTHER INFORMATION

- www.cognitionresearch.org/sp.htm .
- Article: “Big data and the SP theory of intelligence”, J G Wolff, *IEEE Access*, 2, 301-315, 2014.
- Contact:
  - jgw@cognitionresearch.org,
  - +44 (0) 1248 712962,
  - +44 (0) 7746 290775.