

# Notes

*Dr Greig Russell*

*30 June, 2019*

## Contents

<b>1</b>	<b>Synopsis</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>3</b>
<b>3</b>	<b>The crucible</b>	<b>3</b>
3.1	The hardware . . . . .	3
3.2	The software . . . . .	4
3.3	The Operating System . . . . .	5
<b>4</b>	<b>The rise and rise of data</b>	<b>7</b>
<b>5</b>	<b>The ontology of information</b>	<b>8</b>
5.1	Dretske . . . . .	8
5.2	Carnap and Bar Hillel . . . . .	8
<b>6</b>	<b>The lost counterfactual</b>	<b>8</b>
6.1	Lewis . . . . .	8
<b>7</b>	<b>Conclusion</b>	<b>8</b>

## List of Figures

1	Von Neumann Architecture . . . . .	4
2	The essential modern computer . . . . .	6

# 1 Synopsis

Debates on the nature of information continue despite living in the information age (Castells [1997]). There are a plethora of terms such as data, big data, data lakes, machine learning, artificial intelligence, news, and fake news which is, apparently, different from just lies.

Considered will be successive theories regarding the ontology of information before discussing the idea that the key to understanding the ontology of information is that unlike data, information can be used to build “explanatory” theories. Using the perspective of Lewis, such theories counterfactually can provide a sufficient, at least to the observer, causal explanation for some observed phenomena (Lewis [2000]).

Information as a word is an abstract noun. It is not a physical object. Instead, information is an abstraction of some key or vital essence of something else. Information is composed of data, but with something else as well. It is the nature of this something else that appears to cause the difficulties.

For Turing, the father of modern computing, information is held within the algorithm and not within the data (Turing [1937]). Consider a word processor. The validity of the algorithm is based only on whether it performs the correct transformation of the data to the satisfaction of the expert user. This position is utterly reasonable given that a computer was viewed as only a device to support a positivist research paradigm, run by experts. For Turing then, information is the data plus the independent algorithm that describes the causal relationship between the data elements to an expert.

Shannon, as a communications engineer working for Bell laboratories, views information as a measure of surprisal in a message, plus the raw message data (Shannon [1948]). An example is a set of traffic lights at an intersection. Only the first two cars can see the traffic lights and an unfortunate tree blocks the third car’s view. For the first two cars, the light turns green (the data) and that it is now their turn to go (the surprise factor or information). The third car, instead, learns that it is now time to go (the surprise factor) from the other moving cars ahead. When the traffic light becomes visible, the driver receives the data that the light is now green, but no actual new or surprising information. Critical elements for Shannon’s communication theory is that the possible information in a message is the sum of relative probabilities of sub-components that comprise the message. Expressly excluded is any consideration of semantic content within the message itself.

The preliminary argument for this thesis is that if data is surprising, then theories are reassuring, while information is data that is capable of forming a theory. In the same sense, wheat, like data, has potential but is not sufficient to make products for humans. Equally, with grinding into flour, its potential application is spread across many end products such as bread or cakes, of varying quality or relevance.

Data is like wheat, in that it is plentiful and there are complexities around its collection, storage and distribution, which is of no impact on its end users. These atomic data elements joined together, make the elements or the information components that are capable of forming or supporting a theory. Where a theory, reassuring, making sense of these surprising or even frightening information elements as it generally allows for the world to be more predictable for the individual.

In contrast to Dretske [1981] and his concept of “nuclear knowledge,” where such knowledge comprises theories that have been rigorously tested and are hence accepted by all, this conceptualization argues that theories should be ubiquitous. Many such theories will be of poor quality or be merely wrong, but they all share the property of being reassuring to the holder. A trivial example would be the statement;

“I am late for work, but as I know a shortcut, I will still arrive on time.”

So the fright is the running late for work, the reassuring theory is that a quicker route exists, hence the prediction of a positive outcome. Is this theory generalizable, of course not, it covers only one shortcut to one destination. Can one generate a prediction, probably not, but it is reassuring.

This reassuring theory comprises information about route options derived from a raw data source, such as a map. Information in this instance is a collection of data elements (streets), capable of forming a route because they are adjacent to make a theory, or route, as to the time it will take to arrive at the destination.

## 2 Introduction

Despite living in the information age [Castells, 2010], the truth seems to be in short supply despite computer systems flooding us with information or at least data. Currently, terms like “fake news” or claims of state actors distorting elections through false social media claims are prevalent. If the computer is supposed to solve the world’s problems, why can it not tell us the truth?

Partly this is a misunderstanding of the nature of the computer. Despite the exponential growth in component complexity and capacity with a similar accompanying reduction in price, the core concepts have not changed from Alan Turing’s first description [Turing, 1937]. Equally, the basic design has only gently evolved since Von Neumann’s original design for the EDVAC [Von Neumann, 1993].

The computer is a tool to support experts by undertaking complex, large or repetitive transformations of data as described by the control program. The “truth” value of a computer is about whether the right transformation was selected or correctly encoded by the expert. A word processor transforms the text when the user selects options available. These functions could be to highlight a portion of text or build an index to the manuscript. That a program ran successfully means only it fulfilled the exit criteria. Presumably, by creating the required index, but not necessarily if the algorithm had a software bug. Indeed, success does not even imply the truthfulness of the underlying material.

With time computers supported more research programmes and with it, the transfer of data between the different computers used by individual team members became a necessity [Berners-Lee et al., 1994]. Underpinning this communication of data was Claude Shannon’s theory of communication [Shannon, 1948]. Shannon was an engineer with Bell, who was a telephone company focusing on long distance calls or telegrams.

Shannon was clear that his theory of communication did not comment on the truthfulness of content; others disagreed [Weaver, 1953].

## 3 The crucible

In information science, there is a constant dance between the applied and theoretical branches of the discipline. The work of one underpins the next advance in the other. While electronic computers were created to meet human needs when those needs exceeded the capacity of individual humans, information science preceded and enabled their development. The focus for advancing the capacity of computers or their application to solving human problems ever since has been about developing the hardware and software together with theoretical paradigms. To understand information there requires an understanding of computers.

Simply put, a computer is the combination of its hardware and software, and its current state of development is an iterative development from the original concepts developed during and before WW2.

### 3.1 The hardware

In 1945, Von Neumann [1993] first described the components of what he termed “very high speed automatic digital computing system” (section 1.1), and it has become the standard architecture of the computer. The concept has not changed since, even if the component level details have changed dramatically. The aim of this computer, called the EDVAC, was to solve non-linear partial differential equations [Von Neumann [1993], section 1.2).

The current unified concept of a Central Processing Unit (CPU) served two functions. The first was an arithmetic processor to complete basic mathematical operations. For @Von Neumann [1993] this also included more complex functions such as sin, cos, log and square root functions. (section 2.2). The second component was the logic controller (section 2.3). The logic controller was responsible for the orderly execution of a specific set of instructions. The central logic controller was responsible for the orderly operation of the computer as a whole, including hardware components, regardless of any specific program that it was running. It was this

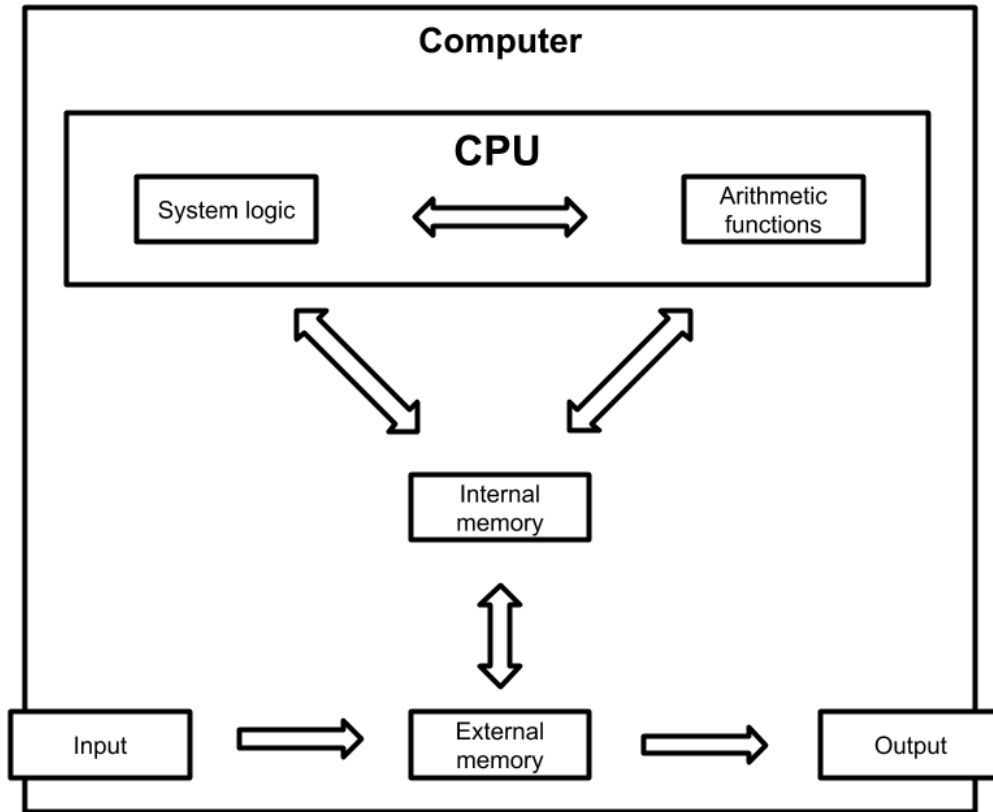


Figure 1: Von Neumann Architecture

latter feature that made this the architecture for a general computer, not dedicated to a specific task like some of the earlier machines (such as the “Colossus” of Bentley Park and the WW2 code breakers fame).

Key to the new design for the EDVAC was the provision of “considerable” memory (section 2.4). Although hilariously small at 34KB by today’s standards, such memory allows for the solution to be broken into components for efficient problem solving within a supervising logic tree. It also allowed for interim states of variables to be available between components to be saved [Von Neumann [1993], section 2.4). Von Neumann [1993] also saw that such generalised computers could be useful for statisticians in regards to the managing of large data samples.

The computer will require input devices [Von Neumann [1993], section 2.6), such as punch cards, and recording this information by use by the computer. Note, for this architecture, the recorded information is not available for the computer. Hence, the computer needs a means of moving the data into the working memory or CPU (section 2.7). Equally, a means of moving the results back to be recorded (section 2.8) and then outputting the results in human-readable forms, such as on a teletype (section 2.9)

### 3.2 The software

Alan Turing outlined the fundamental functional concepts of a computer in his landmark 1937 paper “On computable numbers, with an application to the Entscheidungsproblem” [Turing, 1937]. The EDVAC was supposed to enable such functions to occur within a general computer, as Turing described.

Turing's computer was not the focus of his paper. Instead, it was a thought experiment that served as a vehicle to address the real issue the Entscheidungsproblem or "halting" problem. Yet, this description has served as the archetype for what a computer needs to have to be a complete general computer. The term "Turing complete" has been coined to describe any such computer (or language on modern computers with the abstraction of the programming logic from the hardware into the software) that can perform all the functions of a Turing machine.

The components of a Turing machine are the machine itself and paper tape. The machine consists of a reader/writer head and a set of instructions that depending on the initial state, describe what will happen next. The paper tape consists of a series of symbols, and in Turing's day this was always numbers but what they are is irrelevant to the operation of Turing's machine. The machine reads a symbol on the paper tape, and then performs the specified action by the instruction set based on that symbol. This action may be to change the symbol on the tape or move to another place on the tape or even do both [Turing, 1937].

The paper tape describes the initial state of such a machine. After the machine halts, the paper tape now describes the output of the machine. Turing's research was about whether such machines will always halt and he was able to prove that they did not [Turing, 1937].

For everyone else, Turing's machine is the blueprint of a computer, which was brought to life on physical hardware following the design based on Von Neumann's architecture.

### 3.3 The Operating System

Remarkably, this remains the fundamental design of a computer ever since. The only other significant change was in the 1960s IBM is credited with the introduction of the operating system (OS). This innovation overcame the practical problem of the instruction set is unique to every computer chip. Changing the hardware meant rewriting all of the software, which as the scale of instruction sets grew exponentially was utterly impractical.

The OS became an abstraction that took standardised requests from an instruction set and translated them to the system used by the underlying hardware. For software developers, they could write the code once, and it would run across different hardware configurations. For hardware engineers, they could rapidly innovate without breaking all the client's software.

The standardised computer has become as shown in figure 2. As an example of such a machine, consider the word processor. A user buys a physical machine, which comes pre-equipped with an OS. From the list of word processors compatible with that OS, the user installs one. Their consequent writing becomes the data or paper tape, and the program is the set of instructions to enable the data to be transformed, but the truth function of this transformation is in the instruction set and not in the data.

On reviewing the code, but not the data, one could see that a particular set of instructions will cause a selected text to be made bold. Regardless of what the truth value of the text is, the truth value of the program is whether the text was made bold or italic.

While the differentiation between the two may seem a pedantic quibble. The assumption of truth that the correct rendering of a blog means that the contents of the blog are equally correct has led to vaccination hesitancy to be included in the WHO top10 threats to world health due to incorrect anti-vaccination propaganda <sup>1</sup>. The truth value of the data or content, is for the user. In the era of Turing and von Neumann, users were expert mathematicians or equivalent.

The rest of computing for the last 50 years has been an arms racing between providers of the hardware, the developers of operating systems, and the developers of software. The capacity and capabilities of individual machines has increased dramatically over time, but the core concepts have not.

What has changed is the falling price of computers so that they are now widely available, meaning users are no longer experts, making interpretation of output more fraught. The rise of the internet or interconnected

---

<sup>1</sup>"Ten threats to global health in 2019, from "https://www.who.int/emergencies/ten-threats-to-global-health-in-2019" downloaded 5.May.2019

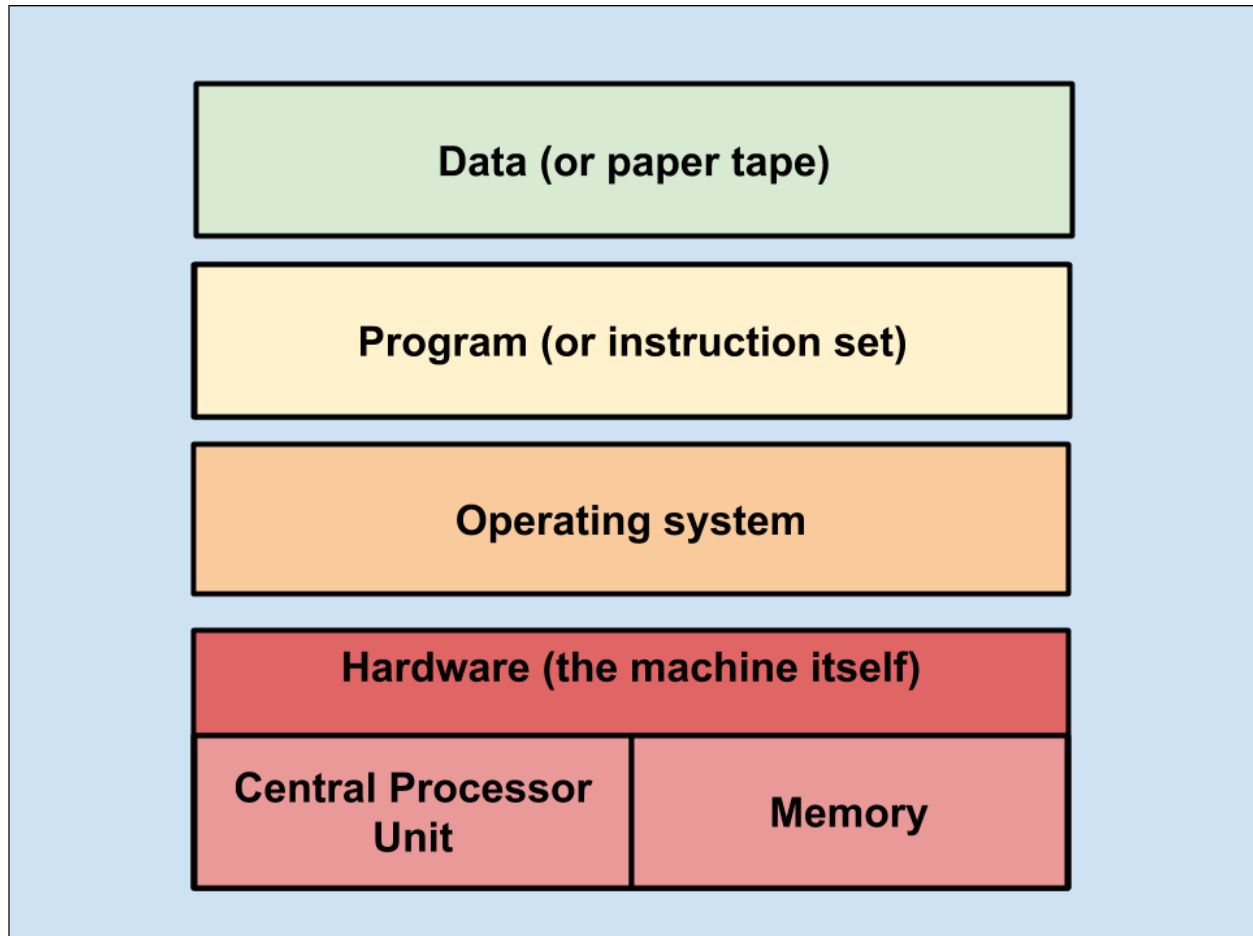


Figure 2: The essential modern computer

computers makes determining the truth at scale even more fraught. Some believe that Shannon’s law or communication protocols has considered this.

## 4 The rise and rise of data

If hardware had been maturing since the 1940s and software has been evolving, then the data has been undergoing a revolution. From being stored on paper punch cards in giant warehouses, and the passive recipient of the software’s transformation, data now is the critical component in the “information age” [Castells, 1997].

Beyond doubt, is the role of Claude Shannon and his mathematical theory of communication [Shannon and Weaver, 1963]. Bell Labs employed Shannon, and the development of his theory of communication occurred within the context of Bell’s business model of focusing on long-distance business communications within the USA.

As Weaver [Shannon and Weaver, 1963, p. 3] stated, the problem of communication comprises three components;

1. The technical problem of how accurately components of the message can be transmitted?
2. The semantic problem of how information can be encoded, a process by definition that is reductionist, yet still retain meaning?
3. The effectiveness problem or rather, how does the information still cause the desired effect in the recipient?

The latter two components are core to Shannon’s view of information and its transmission.

Information, for Shannon, is about the “freedom of choice” between the options one could choose, not the final choice or indeed meaning [Shannon and Weaver, 1963, p. 8].

Consider an example of a baggage handler at an airport. Given several planes, the handler needs to know which particular plane is the destination for this bag. If there are two planes, then a binary system is sufficient. A “yes” or “on” will indicate one plane, a “no” or “off” the other plane. This yes or no status indicator is a single bit of information. The baggage handlers have an index code for each plane (Information) without knowing the contents of the baggage (meaning). Three planes would require a 2-bit information system to differentiate, but would also be able to cope with four different planes ( $2^2 = 4$ ).

In any system, there is a statistical variation between the occurrence of outcomes. Locally, most flights to metropolitan centres and use large planes, less often smaller planes fly to other provincial locations, and occasionally much smaller planes fly to holiday destinations. Shannon and Weaver [1963], therefore, added a probability to the use of each code. The total amount of information (I) generated by event  $s_i$  in a given system is;

$$I(s_i) = -\log_2(p_i) \tag{1}$$

Borrowed from the physical sciences, entropy (H) in this context determines the amount of choice in a system or inverse of the amount of “organised” information in a given system (S) [Shannon and Weaver, 1963, p. 12]. It is the combination of each index code, together with how likely it is to occur.

$$H(S) = -\sum_{i=1}^n p(s_i) \log_2 p(s_i) \tag{2}$$

One can then determine the relative entropy of a given system by comparing its current state to the maximum theretically possible. The redundancy in a given system is  $1 - \text{Relative entropy}$  as this amount of information could be deleted from the overall message, without the information in the message being degraded.

## 5 The ontology of information

### 5.1 Dretske

### 5.2 Carnap and Bar Hillel

## 6 The lost counterfactual

### 6.1 Lewis

## 7 Conclusion

## Bibliography

Tim Berners-Lee, Dimitri Dimitroyannis, A John Mallinckrodt, and Susan McKay. World wide web. *Computers in Physics*, 8(3):298–299, 1994.

Manuel Castells. *End of Millennium: The Information Age: Economy, Society and Culture*. Blackwell Publishers, Inc., 1997.

Manuel Castells. The information age. *Media Studies: A Reader*, 2(7):152, 2010.

Fred Dretske. *Knowledge and the Flow of Information*. reprinted CSLI Publications 1999, 1981.

David Lewis. Causation as influence. *The Journal of Philosophy*, 97(4):182–197, 2000.

Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

Claude Elwood Shannon and Warren Weaver. The mathematical theory of communication–univ. *Illinois press, Urbana, I*, 11:117, 1963.

Alan Mathison Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London mathematical society*, 2(1):230–265, 1937.

John Von Neumann. First draft of a report on the edvac. *IEEE Annals of the History of Computing*, 15(4):27–75, 1993.

Warren Weaver. Recent contributions to the mathematical theory of communication. *ETC: a review of general semantics*, pages 261–281, 1953.