Chapter 1

Information and Quantum Information

1.1 Information is not a thing

Information is ubiquitous. It resides in books on mathematics, plant biology and history. It travels along wires, fibres and through air. It is processed in computers, brains and DNA. It can be stored, deleted and retrieved. But what "is" information? If it were an object – what would be its space-time coordinates? If it were a property – what property would that be for, say, the information revealed by Fermat's last theorem?

The question what information "is", and how it relates to "knowledge", "understanding" and other categories of epistemic character has a long history in pure philosophy, in particular the philosophy of science and the philosophy of language.

In these disciplines information is considered neither a thing nor a property, but rather a kind of semantic code which is carried by messages: *Information is meaning, coded in form.*[1] The string `LOL`, for example, is a message – a string formed from the characters of the standard latin alphabet. A priori such message must not be meaningful. It is only if you are familiar with the cultural habits of the SMS (Short Message Service) that you may be able to decipher `LOL` to mean "Laughing Out Loud!". Depending on the context, you may then find "Laughing Out Loud!" significant or not.

Alices' love to Bob, which she feels very strong about, may be expressed in a poem, first handwritten (a formation of ink on paper), then typed into a computer, where it turns into a sequence of electric pulses, which land on Bobs hard disk as a pattern of magnetic spots, before it is displayed on Bob's screen and makes his heart tumble ...or it could be expressed by a selfie which she takes with her digital camera, the bits and bytes of thich are then transmitted via MMS (Multimedia Messaging Service) to Bobs smart phone where it drowns in the sea of all the other selfies, twits and tweets, and hence goes unnoticed by Bob. In any case there is a certain message which carries information from Alice to Bob.

Tumbling hearts aside, the transmission of information from Alice to Bob would be considered successfull, if the Alices Poem displayed on her screen is identical with the sequence of characters displayed on Bob's screen. A typical question of the information theory would then be the question "How many bits must be transmitted from Alice to Bob so that the poem is faithfully displayed on Bob's screen?" Verbatim (pure ASCII), the poem may come with file-size of 2kB, that is 16000 bits. But that does not mean that the internet will be loaded with 16000 bits for the transmission. It could be less, after all, if she is using a packaging routine first. Or it could be more, because there is noise in the connection line and she must send

[1] Enjoy *Information means meaning in formation.*

twice or even more often.

As the example illustrates the information sciences primary concern is the type of source (natural language vs digital camera), the type of channel (digital electric pulses vs analog em waves), and in particular the amount of raw data (2kB before compression). This peculiar notion of information derives from the needs of telecommunication and computation technologies, which measures information in quantitave terms of resources, efficiency and reliability rather than trying to capture its essence in qualitative terms of meaning or relevance. And since the technologies can not foresee what particular message will be transmitted at any given instance of time, packaging routines and internet protocols are not designed for the most efficient handling of a particular message (like Alice's poem) or even 'worst-case scenarios' (a digital camera file with each pixel a different colour), but rather for the most efficient handling of the typical cases – that is cases the routines and protocols must handle "most of the time" – i.e. with high probability. In short, information sciences is in fact not about information in the sense of meaning or significance, but rather about the characteristics of the variety of sources and channels, and these characterisitics are formulated in statistical terms of probability or likelihood and the like.

The statistcal foundations of the information theory was pioneered by Claude Shannon, who in his 1948 landmark paper first established and subsequently solved two problems, which are at the heart of classical information theory:

(a) How much a partially redundant message can be compressed?

(b) How much redundancy is necessary for reliable communication over a noisy channel?

Here the key concept is redundancy – how frequent the occurrence of a certain

Newtons apple – thought different
The falling apple, which helped Newton to discover gravity, is also a special purpose computer for computing squares. Here is how it works:

: let the apple fall for a time x.

: measure height of fall y.

This is a pretty short algorithm which is universal in the following sense: instead of an Apple, you may decide to drop HAL, an International Business Machine, or any kind of Digital Equipment NeXT to you. You may even decide to drop Texas or the Sun. Note that you can use the same pieces of hardware and equally short an algorithm to compute the square-root of any given number.

character in a given message – and it was Shannon's great insight that entropy provides a suitable measure of redundancy and thus of information.[2]

1.2 Information is physical

Yet communication and computation, the transmssion of information and its manipulation – they all rely on a physical carrier – a sound wave, say, a light pulse, or a magnetization pattern on a hard drive. No carrier – no information. In the words of Rolf Landauer

> Information is physical.[3]

Accordingly all processing of information is a physical process,[4] and sending, transmitting and receiving of information are just actions of preparation, propagation

[2]Again – the information theorist concept of information is based on a purely syntactic notion of redundancy. From the Shannon point of view the message `DON'T SHOOT THE PIANIST!` may well be as informative as the message `?>+89$*898*)9P<(><$8-`. The latter message will most likely be perceived meaningless, and thus may not appear informative, but there is no contradiction. Information theorist notion of information doesn't care about meaning (it doesn't deal with semantics) nor does it care whether a certain piece of information is relevant or not (it doesn't deal with pragmatics) – it only deals with syntax.

[3]It's good a slogan, but bad a definition. "Love is physical!" is certainly true, as "love" may indeed be viewed a specific physical state of arousal, yet that doesn't tell much about what love "is". Love is what the theory of love is about. And the theory of love is poetry.

[4]One could turn this around: every physical process can be considered a process of information processing – see Fig. **??** for an example. The evolution of the entire universe is a gigantic process of information processing for which, however, neither the initial problem nor the solution is known to date. There are indications that the answer is 42, but this leaves unanswered the question what question the answer answers, and how we could possibly verify (or falsify, for that matter) the proposed answer, which is 42.

and measurment performed on these carriers.

The physical nature of information processing is nicely illustrated in Landauer's principle which states that the erasure of one bit of information necessarily increases the entropy of the environment by an amount of at least $k_{\mathrm{B}} \ln 2$, where k_{B} is the Boltzmann constant. From the physics point of view, Landaur's principle is a simple consequence of the second law of thermodynamics. With an environment at temperature T, the erasure of a bit then implies an irreversible transfer of energy, at least $k_{\mathrm{B}} T \ln 2$, from the bit to its environment. So – every time you delete an e-mail from your computer you necessarily contribute to global warming!

Surely enough – for the standard information theory the physical nature of the information carrier is irrelevant. There is no need to distinguish between "accoustic information", "electromagnetic information" or "printed information". When in an everyday telephone conversation the words carried by a sound wave are converted into an electric signal and at the receiver side back to a sound wave, the information travels undisturbed from sender to receiver, and if there are disturbances these are well understood and there are means to correct for them. The abstraction from the physical carrier, which within the validity-range of classical physics is well justified, is of great importance for the success of the standard information sciences as it allows to concentrate on the general principles and concepts rather than getting tangled up in the physics of the various carriers.

Yet with the carriers on the quantum scale, it turn out, that certain elementary processe, like the faithful copying of information, become problematic, while other devices, like the "Square-Root of `NOT` gate", which is a sheer impossibility according to standard informtion theory, become accessible and open a whole new universe for the manipulation and processing of information. In particular, the faithful transferral of information form one carrier to the other – a cornerstone for the standard information theory – is no longer possible if one considers the transferral of information form a quantum carrier to a classical carrier and back to a quantum carrier. There is no way, to read the polarization of a qubit which is in a prior unknown state which would enable you to prepare another qubit in exactly that state.

Thus with the information carriers becoming quantum, a new type of informtion enters the stage, the preliminary definition of which has been given by Reinhard Werner, a leading pioneer of the field,

> Quantum Information is that kind of information, which is carried by quantum systems from the preparation device to the measurung apparatus in a quantum mechanical experiment.[5]

We will see in due course what "that kind" refers to ...

1.3 Complement: Probability in a nutshell

Probility theory provides mathematical models for the description of experiments which are goverened by chance. Any such **statistical experiment** (like flipping a coin, rolling a die, or spinning a wheel) is associated a set Ω, called the **sample space**, each element of which is a possible **outcome** of the experiment. In most of the cases considered here, the sample space is a finite set, in case of *rolling a die* the set of faces $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$. A possible outcome would be ⚂ (the top face of the die after rolling). In case of *flipping a coin*, the sample space is given by $\Omega = \{\mathtt{H}, \mathtt{T}\}$, with `H` a possible outcome (meaning the coin lands "Head up" after

[5]Quoted from *Quantum Information Theory – an Invitation* by Reinhard Werner, arXiv:quant-ph/0101061v1 (2001)

flipping).[6] A typical experiment with an infinite sample space is *spinning a wheel*. The outcome is the total angle, modulo 2π, turned up by the wheel when if finally comes to rest. The sample space is the interval $[0, 2\pi[$, and this is an infinite set.

In contrast to deterministic experiments, where the outcome is fixed by the experimental conditions and "is always the same", in case of statistical experiments, the outcome may vary from trial to trial – the outcome is a **stochastic variable**. Consider an experiment with finite sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$. All what can possibly be said about a particular outcome ω_i is that it will occur with a certain **probability** $p_i := \mathrm{Prob}(x = \omega_i)$ (read: with probability p_i the outcome X assumes the value $x \in \Omega$ given by $x = \omega_i$). A fair die, for example, is characterized $p_i = \frac{1}{6}$ but you may recall, that there are also unfair dice on the market (in which case they are called biased). The only restrictions the p_i must obey are (i) $p_i \geq 0$ (probabilities are non-negative), and (ii) $\sum_i p_i = 1$ (the experiment has an outcome). The assignment of probabilities to the elements of a finite sample space Ω is frequently coded in a function $p : \Omega \to [0, 1]$ with $\sum_{x \in \Omega} p(x) = 1$, called **probability mass function**, i.e. $p_i = p(\omega_i)$.

Any subset of a sample space Ω constitutes an **event**. The subset $A :=$ {⚁,⚃,⚅} $\subseteq \Omega$, for example, constitutes the event "even number of pips". If you roll a die, and the outcome is ⚃, the event A has occured, if the outcome is ⚂ it has not. The **event space**, denoted $\mathcal{B}$, is a collection of all the events one might be interested in. Technically speaking, it is a σ-algebra, that is a system of subsets of Ω, which contains Ω (the "sure event"), and which is closed under complement and under countable unions. The **Kolmogoroff axioms** of probability theory then assign any statistical experiment a **probability space** $(\Omega, \mathcal{B}, P)$ where P stands for **probability measure**, that is a function $P : \mathcal{B} \to [0, 1]$ such that for all $A \in \mathcal{B}$ we have

[6] A statistical experiment with two outcomes "success" and "failure" is sometimes called a **Bernoulli trial**.

$P(A) \geq 0$ (probabilites should connect to *relative frequencies* – see below), for disjoint sets $A \cap B = \emptyset$ we have $P(A \cup B) = P(A) + P(B)$ (exclusive-or is additive), and $P(\Omega) = 1$ (the sure event occurs with certainty).[7] For experiments with a **finite sample space** Ω, like flipping a coin or rolling a die, the experiment can always be modeled with an event space $\mathcal{B} = 2^{\Omega}$ (the power set of the sample space), and for given probability mass function p, the probability that event A occurs is given by $P(A) = \sum_{x \in A} p(x)$. For experiments with an infinite sample space things are a little bit more complicated. Spinning a wheel, for example, comes with an event space which consisits of all intervals $[a, b] \subseteq [0, 2\pi[$, their countable unions and intersections. The reason for this construction is pretty simple: spinning an ordinary wheel, you will never end up with an angle exactly $x = \pi$ (you will never encounter the event $\{\pi\}$), although $x = \pi$ is a possible outome. However the probability for the event that the wheel's outcome falls into a particular interval $[a, b]$ is finite, for balanced a wheel given by $P([a, b]) = \frac{|b-a|}{2\pi}$.

The number $P(A \cap B)$, also written $P(A\&B)$, called **joint probability**, measures the probability that the outcome of an experiment realizes jointly both events A and B. In rolling a fair die, for example, the probability that the outcome realizes the event $A =$ "even number of pips" and $B =$ "more than three pips" is given by $P(A \cap B) = \frac{2}{3}$. If $P(A \cap B) = 0$ the events A and B are said to be **statistically exclusive**,[8] and if $P(A \cap B) = P(A)P(B)$ they are said to be **statistically independet**. For the fair die, for example, the events $A =$ "even number of pips" and $B =$ "more than four pips" are independent.

In case that B comes with a non-zero probability, $P(B) > 0$, the joint probability

[7] In the exercise you will learn, that $\mathcal{B}$ necessarily also includes the empty set $\emptyset$ (the "impossible event") with $P(\emptyset) = 0$ (the impossible event never occurs).

[8] This is weaker than $A \cap B = \emptyset$.

can be factorized

$$P(A|B) := \frac{P(A \cap B)}{P(B)} . \tag{1.1}$$

In the exercise you learn, that for given B the function $Q(A) = P(A|B)$ indeed defines a probability measure on $\mathcal{B}$. It is called **conditional probability**, interpreted "the probability that A occurs, given B".

1.4 Complement: Physics and Probability

A **physical experiment** is a repeatable procedure, described in ordinary language in a lab-manual and/or physics text-book, which consists of a *preparation procedure* and a *measurement procedure*, both of which are executed on a certain *entity* S (an electron, a die, coin, wheel, or potatoe, but not the whole universe, for that matter). Suppose that S can assume a only a *finite* set of possible configurations Ω. The **preparation procedur** is then coded in a probability mass function $p : \Omega \mapsto [0, 1]$, where $p(x)$ is the probability, that the preparation procedure prepares the system in configuration $x \in \Omega$. In the physics mumbo-jumbo the pair (Ω, p) defines an **ensemble** with Ω the **phase space** and p the **state** of the system.[9]

The **measurement procedure** is typically coded in a real-valued function $f : \Omega \to \mathbb{R}$, with $f(x)$ the value displayed by a f-measurement device if it encounters the system in configuration x. With the outcome X of the preparation procedure being stochastic (X takes the value x with probability $p(x)$), the outcome Y of the measurement procedure is also a stochastic variable (Y takes the value $y = f(x)$ if X takes the

[9] The set of possible configurations of a free particle in one dimension is the phase space $\Gamma = \{(q, p)|q \in \mathbb{R}, p \in \mathbb{R}\}$ with q the position and p the momentum. This space is infinite.

value x). The **expectation value** of f is defined[10]

$$\langle f \rangle_p = \sum_{x \in \Omega} p(x) f(x) \,. \tag{1.2}$$

where the subscript is a reminder for the preparation p.

Probabilities finds a natural interpretation in terms of so-called **relative frequencies**. Rolling a die N times, and counting a particular property $A \subseteq \Omega$ a "hit", the data collected in the n^{th} run of the experiment (a run consists of N rolls) yields a number (relative frequency of hits)

$$r(A, N, n) = \frac{\text{Number of hits on } A \text{ in } n^{\mathrm{th}} \text{ run}}{N} \,. \tag{1.3}$$

It is the great promise of nature that – for properly executed experiments – in the limit $N \to \infty$, the relative frequencies $r(A, N, n)$ assume a n-independent – that is "objective" – value *with certainty*,

$$\lim_{N \to \infty} r(A; N, n) = P(A) \,. \tag{1.4}$$

We believe in nature, but we must mention that the interpretation "probabilities are relative frequencies" has quite some philosophical hazards. "With certainty" means "with probability arbitrarily close to one", and thus the argument which defines probabilities via relative frequencies is circular.

Along the same lines – strictly speaking, probabilities can neither be verified nor falsified. The promise "this die is fair" claims, among others, $p(⚄) = 1/6$. If you are unlucky (or lucky?), and each of your trials turns up ⚄, you would be tempted to claim that the promise has been falsified. Yet it has not. You did not, after all,

[10]The angular brackets $\langle \cdots \rangle$ are sometimes written $\mathsf{E}[\cdot]$ as mnemonic for "expectation".

roll the die infinitely often. On the other hand – after analyzing very long sequences up to $N = 10^{43}$ – you find $r \approx 1/6$ for all faces, you have no proof either: the die could nevertheless be biased, despite the promise of the contrary (there could be an internal mechanism in the die such that after 10^{43} rolls, the die always comes up ⚅). With the empirical data giving no conclusive answer to the question "fair or not fair?", probabilistic propositions are contingent for the facts, that is we are *not forced* to apply probability theory, but we rather *decide* to apply probability theory. As with any decision, this may turn out to be a bad decision, but fortunately, depending on our frustration with the experimental data confronting our statistical model, we may well revise our decision and look for another model. For the quantum mechanics, by the way, no better than the probabilistic modeling has been devised so far.

In the laboratory, then, one does not aim to **verify** (or **falsify**) a propabilistic statement, but one rather seeks to **test** it. These tests are done by running the same **physical experiment** over and over again, take note of the outcome of each trial, and derive their relative frequencies. But such test a can only be reasonbly confronted with a theory if its results – that is the relative frequencies – are in some sense reproducible. Thus

> Probability theory applies, whenever the same physical experiment produces different outcomes, but with reproducible frequencies.

Here the critical term is "the same physical experiment". What makes two physical experiments "the same" if the experiments may come with different outcomes? It certainly does not mean "using the same photon again". A photon, after it has been detected, is gone, and simply can not be used again. Also, the evocation of "many copies" of the photon is not helpful, since at this point it is not clear what constitutes a faithful copy. Instead of getting tangled up in a philosophical debate about “sameness”, “identity” and the like, we here assume a pragmatic attitude and simply declare

> two physical experiments to be the same if they deal with the same sort of system undergoing the same procedures for the preparation and the same procedures for the subsequent measurement.

This postulate links the abstract notion of a statistical experiment with conrete actions in the laboratory, but does not exorcise the demon which comes with the probabilistic paradigm: quantum mechanics no longer refers to “the concrete electron Fritz, which I hold in my right hand”, but refers only to an ensemble of identically prepared electrons. Notions like “collaps of wavefunction” do not refer to an individual particle, but only to an abstract entity “ensemble”. With the 1925 quantum revolution, the subject of physics is no longer the concrete individual electron, potato or universe, but the abstract ensemble of identically prepared electrons, potatoes or universes.

A very useful class of measurement devices are detectors. A **detector** is a device which, on a given outcome, either “clicks” or does nothing (does'nt click). Mathematically, such device a is described by a function

$$f \ : \ \Omega \to [0,1] \tag{1.5}$$

$$x \mapsto f(x) \tag{1.6}$$

with interpretation “$f(x)$ is the probability(!) to click given x is the outcome”. Since the outcome is unknown, the expectation value $\langle f \rangle_p \equiv \sum_x p(x) f(x)$ gives nothing but the probability, that the detector in fact clicks, $\langle f \rangle_p = \text{Prob}(\text{“click''})$. This interpretation is consistent as the detectors inequality $0 \le f \le 1$ immediately implies $0 \le \langle f \rangle \le 1$ as it should be. I order to see that the probability interpreation really makes sense we consider two simple examples.

The first example is a **perfect detector** for a given event A. Mathematically, such detector is described by the **characteristic function**

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases} \tag{1.7}$$

Consider for example a perfect detector for the event $A = \{$⚄$\}$. In this case we immediately obtain $\langle \chi_A \rangle = p($⚄$)$ as ist should be: a preparation (ensemble) in which ⚄ is prepared with probability $p($⚄$)$, a perfect detector which asks "has ⚄ been prepared?" should respond "Yes!" (via "Click") with probability $p($⚄$)$. In particular, if ⚄ has been prepared for sure, i.e. $p($⚄$) = 1$, we want to – and indeed do – hear a click every time the die is measured, $\langle \chi_A \rangle = 1$. Perfect detectors are quite important – so how do we reckognize them? Mathemtaically, the answer is exquisitely simple: f describes a perfect detector if and only if it is **projection valued**, $f^2 = f$. If only all were so simple …

The second example is an **imperfect detector** for the event $A = \{$⚄$\}$, say. An imperfect detector is a detector (a "clicker") which (1) never clicks if the outcome differs from ⚄, but (2) not always clicks if the outcome is in fact ⚄. Quantifying the perfection by a number η, with $0 \le \eta \le 1$, the mathematical model of our detector is the function $f = \eta \chi_A$. Indeed, the mean value $\langle f \rangle = \eta p($⚄$)$ is nothing but the probability that ⚄ was actually prepared AND the detector was triggered. Surely, if ⚄ is produced with certainty, $p($⚄$) = 1$, the detector will only click with probability η. Good detectors have η close to one (called "unit quantum efficiency"), bad detectors have η close to zero. We should mention that realistic detectors are imperfect detectors with a slightly different meaning of imperfection: they are detectors which not always click if they see a ⚄ (non-unit quantum efficiency), but sometimes even click if they don't see a ⚄ (so called dark counts). Exercise: try to make a statistical model for such a detector. Why do you have the feeling that our approach could not cover such a detector? Hint: Consider the case that actually no die is send to be measured, but still your detector clicks from time to time ... Contemplate where this counts could come from (where the energy could come from), and then realize that – using a much broader interpreation – our modeling in fact could cover even such a detector (what about using "no die put in" as a value of a stochastic variable "existence").

The next type of detectors, which is slightly more versatile than the simple detector described in the previous section, is a **multi channel detector**. A multi channel detector is specified by a set of detectors $f_i \geq 0$, $i = 1, \ldots, N$, with $\sum_i^N f_i = 1$ (one click for sure). The interpreation is $\langle f_i \rangle_p = \text{Prob}(\text{"click" in channel } i)$. Noteworthy, in a multi channel detector the total number of possible channels, N, is not restricted in general: it may be less or even more than the total number of elementary events $|\Omega|$. Consider the die. Recall $\Omega = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$ the sure event. A possible, yet completely useless $N = 10$-channel detector, is specified by $f_i = \frac{1}{10}\chi_\Omega$. Given this kind of detector, you will observe the ith channel to click with relative frequency 1/10 *independent of the die's preparation*. Truly useless indeed – you can't infer anything about the preparation from any observed sequence of clicks. But very useful turned the other way round: having such a device you could build a perfect random number generator! Exercise: Build a most useful perfect random number generator from a most useless imperfect detector.

Finally the meters – the measurement devices you are so familiar with from "analog" physics. A **meter** is a measurement device which – in the most general setting – we may describe by a function

$$f \; : \; \Omega \to \mathbb{R} \tag{1.8}$$

$$x \mapsto f(x) = \sum_{i=1}^{N} m_i f_i(x) \tag{1.9}$$

with f_i a detector observable, $0 \leq f_i \leq 1$, $i = 1, \ldots, N$, and m_i the meter's readout value in channel i.

Having glimpsed at the zoo of measurement devices (there are many more!), lets apply Ockham's razor. Here is how it goes. There are only perfect detectors. Imperfect detectors are the concatenation of a noisy channel and – downstream – a perfect detector for the channel output.

1924	H. Nyquist	First attempt to measure information
1927	R.V.L. Hartley	Expands on Nyquist, introduces logarithm
1948	C. Shannon	Expands on Hartley's ideas in *A Mathematical Theory of Communication*; foundation of "Information Theory".
1948	N. Wiener	Same as Shannon in *Cybernetics – Control and Communication in the Animal and the Machine*.
1961	R. Landauer	Erasure of one bit generates heat $\Delta = k_\mathrm{B}T\ln 2$ (Landauer's principle). Landauer's principle can be used to exorcise Maxwell's Demon (Maxwell 1871).
1971	C.F.v. Weizsäcker	Introduces the "Ur" (nowadays called qubit) as quantum elementary carrier of information.
70's–80's	many	Reversible computing (Bennett), quantum mechanical computation (Feynman, Benioff), quantum Turing machine (Deutsch), ..., quantum cryptography (Bennett et al).
90's	many more	quantum algorithm for fast factoring (Shor 1994), scheme of an ion chain quantum computer (Cirac and Zoller 1995), quantum algorithm for fast database search (Grover 1996), demonstration of long-distance quantum key distribution, quantum teleportation.
2000+		commercial quantum key distribution, demonstration of few qubit quantum computer ...

Figure 1.1: A brief history of the physics of information

