Quantisation and Prediction

Another look at the aim and structure of quantum theory

Jean-Michel Delhôtel

London School of Economics and Political Science

Department of Philosophy, Logic and Scientific Method

Thesis submitted for the Degree of Doctor of Philosophy

University of London

2004

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Our present quantum mechanical formalism is a peculiar mixture describing in part laws of Nature, in part incomplete human information about Nature – all scrambled up together by Bohr into an omelette that nobody has seen how to unscramble. Yet we think the unscrambling is a prerequisite for any further advance in basic physical theory...

E.T. Jaynes

Erwin with his psi can do
Calculations quite a few.
But one thing has not been seen
Just what psi really mean.

Felix Bloch

There exists a body of exact mathematical laws, but these cannot be interpreted as expressing simple relationships between objects existing in space and time.

Werner Heisenberg

...if quantum theory were not successful pragmatically, we would have no interest in its interpretation. It is precisely because of the enormous success of the QM mathematical formalism that it becomes crucially important to learn what that mathematics means. To find a rational physical interpretation of the QM formalism ought to be considered the top priority research problem of theoretical physics; until this is accomplished, all other theoretical results can only be provisional and temporary.

E.T. Jaynes

ABSTRACT

It is argued (Part A) that quantum mechanics can be derived as a principle-based dynamical framework, the basic equation of which is an alternative form of the Hamilton-Jacobi equation. Schrödinger's equation obtains as a result of linearising that equation, and so-called wave functions can be given no straightforward physical interpretation. It is suggested, partly in relation to a theorem by Gromov, that a finite action quantum would make it practically inevitable, for purposes of prediction, to resort to a probabilistic formulation. The structure of the space of square-integrable solutions of the Schrödinger equation happens to lend itself to the introduction of the appropriate kind of predictive scheme. Investigating the nature and scope of such a scheme is the subject of Part B. It is shown that basic features of the formalism of quantum theory, like composition rules for 'amplitudes' or the 'Born' probability rule, can be derived independently of any physical assumptions. A generalisation of the basic formalism using tensor product composition appears to be required if all correlations are to be extracted from locally accessed data. A detailed discussion of quantum teleportation leads to the conclusion that a 'one-shot' account leads to a distorted picture of what is actually achieved. An analogy with classical cryptography is made and the statistical significance of the 'transfer', which does not require introducing any novel form of 'quantum information', is emphasised. Results obtained over the last decade using the extended formalism of positive operator-valued measures are reviewed and discussed. These lend further support to the idea that the set of basic 'quantum' rules functions as a general kind of probabilistic scheme for prediction, the structural features of which are not constrained in any direct way by the underlying physics. On the other hand, the very existence of such a predictive framework hinges on selecting a particular class of solutions of the Schrödinger equation, which selection has been incorrectly interpreted as reflecting a physical necessity.

Contents

Introduction

Three quarters of a century have now gone by since Bohr's Como lecture (1927). All the same, there is as yet no consensus regarding how the most basic features of the quantum-mechanical formalism should properly be understood. Half seriously, one could say that quantum physics has been a tremendously successful answer to a question that we have yet to ask and formulate clearly. Up to the present day, the dominant trend has consisted in defending one's favourite interpretation against other major competitors or in the face of a number of 'no go' theorems. The feeling is also widespread that the most counterintuitive aspects of quantum theory point to an inherent weirdness of the physical world, involving uncontrollable influences at a distance, a blurring of the distinction between being and non-being or some inscrutable variety of holism. Indeed, the popularity of baffling interpretations like Many Worlds and its variants testifies to the fact that (too) many physicists or philosophers readily surrender to the idea of a sweeping 'non-classicality' of quantum mechanics. By contrast, comparatively rare attempts to ground the theory on a small set of well motivated, and perhaps more mundane principles having clear physical significance, tend to be considered futile or regressive ('hidden variables').

Whilst 'no go' proofs can be technically impressive, they all boil down to the acknowledgment of two well-known and related basic features of the quantum-mechanical formalism: a metric vector space structure and a non-commutative algebra of linear operators. These underlie both the 'non-classicality' and the efficiency of the algorithm whereby expectation values and probabilities are routinely calculated in the laboratory. The rise of Quantum Information Technology (QIT) has recently prompted a revival of interest in the logico-algebraic structure of quantum theory, which had once been the rather exclusive concern of a small band of quantum logicians and philosophers. New 'hard' results, admittedly of unequal conceptual importance, have been obtained over the last ten years. Technical breakthroughs like Shor's algorithm for factoring large numbers into primes or technological prospects like that of manufacturing a quantum computing network in the not-too-distant future have bred optimism. Almost inevitably, this has led some to hope that a change of focus toward information-theoretical concepts and methods might hold a (*the*?) key to a final, if belated understanding of the aim and structure of quantum theory. In all fairness, nothing in what has been achieved in QIT so far, or what can be reasonably

foreseen, really leads one to suspect that any such revelation (revolution?) is under way. On the other hand, the rejuvenation of old concerns, the fresh outlook of new participants in the debate and – last but not least – the sharpening and more extensive use of some relatively new tools like positive operator-valued measures might all contribute in a beneficial way to our understanding of quantum theory. Thus, Chapter 9 of this thesis is devoted to some recent and outstanding contributions by theorists who all participate to some degree in the QIT enterprise. The upshot is a re-evaluation of the nature and purpose of the basic formalism of quantum theory: this is seen to function as a linear probabilistic scheme for prediction, the mathematical structure of which is essentially independent of any assumption regarding the nature or behaviour of physical objects.

It is one of the aims of this dissertation to convince the reader that the whole of quantum physics can justifiably be established as following from the satisfaction of a fundamental structural principle, in a similar sense as the refoundation of mechanics prompted by the Special Theory of Relativity (STR) can be shown to follow from assuming the universal validity of a single[1] relativity principle. Another is to show how such a well-motivated principle view of 'quantum' dynamics (*Part A*) can be maintained consistently with the idea that the basic formalism of quantum theory, as it is currently used and debated upon, essentially *functions as* a linear predictive scheme (*Part B*). Indeed, a driving conviction behind this work is that no progress will be made in our understanding of quantum mechanics unless what pertains to the theory as an alternative to classical mechanics is clearly distinguished and properly separated from those features of the formalism that make all the difference between 'quantum' evaluations of probabilities and ordinary ('classical') ones. Claiming that such a distinction is both meaningful and necessary stands in contrast to widespread views whereby, for example, the squared modulus rule for the calculation of probabilities would directly reflect the physical nature and radically non-classical features of putative 'quantum states'.

Part A opens (Chapter 1) with a historical outline. The intention is here to help set the contents of the next two chapters in proper perspective. Although the likely relevance of action-related aspects of dynamics to the development of a quantum mechanics was recognised very early, it has in our view subsequently been unduly neglected as a central subject matter in the debate about the theory. Section 2.1 reviews the attitude of some of the leading pioneers of atomic physics in regard to another key issue: that of whether 'wave functions' or Hilbert space vectors in quantum theory should be interpreted as representatives of *states* of physical systems (the representation of physical quantities within a Hilbert space framework is

[1] See Section 3.1.

addressed in Section 2.2). Following the lesson of STR (Section 3.1), it is argued that quantum mechanics might, on account of its status as a fundamental framework for theoretical physics, best be formulated as a *principle theory*. After considering what is to be required of a principle view of quantum mechanics (Section 3.2), a recent (2000) and compelling derivation of 'quantum' dynamics from a single principle is presented in some detail (Section 3.3). Such a principle-based refoundation of quantum mechanics sheds light on the fundamental nature and scope of the standard quantum-mechanical framework, based on the Schrödinger equation or on the Heisenberg picture. It is a remarkable fact that, although the suggested refoundation of quantum mechanics involves no randomness or lack of determinism, square integrable solutions of the Schrödinger equation have, nonetheless, just the requisite properties for setting up a novel kind of probabilistic framework for prevision. *Part B* is devoted to investigating prerequisites of such a scheme, its basic structural features and predictive capacity. In Chapter 4, a statement of the standard 'rules' of quantum theory is followed by a brief discussion of the difficulties raised by the common urge to reify or to objectify 'state vectors' or density operators. Derivations of elementary features of the formalism – the composition rules for amplitudes and the modulus-squared probability rule – are examined in Chapter 5. These lead us to the suggestion that quantum theory should be conceived as supplying a general-purpose linear scheme for prevision – outlined in Section 6.1 – whose characteristic structure essentially reflects consistency constraints. Section 6.2 addresses the question of Hilbert space angle as an appropriate measure of statistical overlap within such a scheme, and Chapter 6 ends with a discussion of the controversial 'reduction' rule..

Tensor product composition allows the predictive framework sketched out in the previous chapter to be applied to 'fragmented' preparations (Chapter 7). Comparison of locally gathered data may then reveal correlations that cannot be accounted for in terms of simple-minded common cause explanations or preordained 'instruction sets'. Whether this fact warrants considering 'spooky' influences at a distance or other forms of 'communication' between causally disconnected (sub)systems is addressed in the form of a case study (Chapter 8). Quantum teleportation seems to exemplify in a striking manner some of the most baffling implications of quantum theory: potentially measurable attributes of a system, embodied in its 'state', appear to be transferable to a distant system although sender and receiver lack the information that would appear to be needed for any such transfer to be achieved. The discussion of Chapter 8 emphasises the fundamentally statistical significance of the 'transfer'. In contrast, a literal 'one-shot' reading of the transformations involved gives rise to incorrect expectations and claims that are empirically vacuous. Besides being quite insufficiently motivated, the idea of a backwards-in-time

propagated 'quantum information' transfer between two or more subsystems is also shown to be very implausible. A parallel is made with classical cryptography, and it is seen that quantum teleportation, dense coding or entanglement swapping lose much of their mystery provided conditions and procedures for extracting and comparing the relevant data are properly accounted for.

Further support to a view of standard quantum theory as a linear predictive scheme can be adduced (Chapter 9) using the resources of positive operator-valued measures (POVM). For example, the rule for updating a density operator assignment in the light of information obtained through measurement can be given a form that is more similar to classical Bayesian update than the reduction rule affords. Section 9.3 reviews and discusses a new derivation of the basic structure of quantum theory (in a generalised POVM version) from a set of just five axioms. It is very significant that the statistical algorithm of quantum theory should be found to differ from a linear representation of the classical probability calculus by only one axiom, and that none of the five axioms depends on assumptions regarding the nature or properties of physical systems.

Part A

1

The Emergence of quantum mechanics
A historical outline*

Toward the end of the nineteenth century, basic tenets of classical (Newton-Maxwell) physics were unexpectedly challenged. Whilst the puzzle of the invariance of the velocity of light *in vacuo* found a rather early resolution (1905), more than twenty-five years would pass before the emergence of a comprehensive framework in which the remaining issues would (seem to) be settled. The rise of quantum mechanics could have benefited from a far more auspicious start[1] than a complicated state of affairs at the crossroads of electro- and thermodynamics. It was the urgency of resolving what was rightly perceived as an acute difficulty – the black-body problem – that prompted the first step toward the emergence of quantum theory.

Kirchhoff (1859) had showed that the radiation inside an isothermal enclosure (cavity) would behave like an absorber of all incident radiation, or 'black body' radiator, at the same temperature. Only in 1895 could such an enclosure be constructed (Lummer and Wien) and the empirical study of the spectrum of black-body radiation commence, by which time Wien's 'displacement law' (1894) had been formulated: the emissive power E_λ per unit interval of wavelength is inversely proportional to the fifth power of the wavelength λ and proportional to a single function $\varphi(\lambda,T)$ of the wavelength and temperature. The maximum value of E_λ occurs at a wavelength that is inversely proportional to the absolute temperature of the black body. To account theoretically for the empirically observed distribution of E_λ curves would then be tantamount to working out the expression of $\varphi(\lambda,T)$, which was to be independent of the particular properties of any material substance involved. As more and more desperate attempts were made to match the form of data curves, reasonable optimism quickly gave way to a crisis situation. The Rayleigh-Jeans law[2], derived in 1900 through the application to cavity radiation of the equipartition theorem, agreed with the data in the low frequency regime where Wien's

---

* This chapter is greatly indebted to Jammer 1974,1989 and to Mehra and Rechenberg 1987.

[1] The issue of quantisation could conceivably have arisen in the context of the classical kinetic theory (see Mott 1964).

[2] According to the Rayleigh-Jeans law, the energy density $u(\nu)$ is proportional to the temperature and to the square of the frequency ν ; more precisely, $u(\nu) = \frac{8\pi\nu^2 kT}{c^3}$.

foundered. On the other hand, not only did the Rayleigh-Jeans law not overlap with Wien's where the latter matched the observations, but it also implied that the energy density should shoot up without limit as frequencies become arbitrarily large, in clear contradiction with both observations and sensible expectations. The predicted divergence ('ultraviolet catrastrophe[1]') was a telling symptom that something was fundamentally amiss in the classical approach to the problem.

Max Planck entered the scene in 1897. What he sought to achieve was to show that, regardless of their initial conditions, an assembly of linear harmonic oscillators of frequency ν would be collectively 'driven' to an energy distribution that corresponds to the observed spectrum of black-body radiation. His approach paralleled, albeit in an electromagnetic rather than molecular setting, Boltzmann's earlier attempt to show that a spontaneous evolution to equilibrium is irreversible (*H* theorem). Eschewing any reference to the equipartition theorem, which would have sent him along the same track as Rayleigh and Jeans, Planck introduced an 'entropy' function S of the average energy U of a harmonic oscillator at temperature T, satisfying the equation

$$\frac{\partial^2 S}{\partial U^2} = \frac{C}{U}, \qquad (1.1)$$

where C is a constant. Associated with his former derivation of the equilibrium conditions, this equation led to Wien's law. However, the reported failure of that law in the low-frequency and high temperature regimes soon led Planck to revise his initial approach. Since the proportionality to temperature of the energy density was implied both by the empirical data and the Rayleigh-Jeans law, which followed from the equation

$$\frac{\partial^2 S}{\partial U^2} = \frac{C}{U^2}, \qquad (1.2)$$

Planck ventured an interpolation between (1.1) and (1.2), namely

$$\frac{\partial^2 S}{\partial U^2} = \frac{a}{U(U+b)}, \qquad (1.3)$$

where *a* and *b* are constants. This form was precisely cooked up to guarantee that his former equation (1.1) and that (1.2) leading to the Rayleigh-Jeans law would both be recovered in the regions where they were found to be empirically adequate. The ensuing radiation law was found to be in very good agreement with the available data.

Planck's achievement was the outcome of a "lucky interpolation", as he recognised it himself. Looking for a microscopic grounding to the new law, he followed once again Boltzmann's footsteps, defining the entropy S_N of

[1] The expression was coined by Ehrenfest in 1911.

a system of N oscillators of given frequency as the logarithm of the number *W* of complexions i.e. distributions of those oscillators that were compatible with the energy of the system. If that number *W* was to be determined using a combinatorial procedure, then the total energy E of the system had to be regarded as decomposable into a finite number of discrete *Energie-elemente*. Planck made the *ad hoc* hypothesis that an integral number P of discrete and equal energy elements or *quanta* ε were to be distributed among the N resonators of frequency ν, i.e. that $E = P\varepsilon$. Given that assumption, *W* is just the number of ways P energy elements ε can be distributed among the N resonators. It takes but a fairly simple calculation to find an expression for S_N such that the entropy $S = \frac{S_N}{N}$ of a single resonator satisfies Planck's interpolation formula[1] (1.3) (with $U = \frac{E}{N}$). Given the condition $\frac{\partial S}{\partial U} = \frac{1}{T}$, the average energy of the resonators of frequency ν is

$$U = \frac{\varepsilon}{\exp(\frac{\varepsilon}{kT}) - 1}. \qquad (1.4)$$

This formula, obtained in 1900, agrees with Planck's former expression of U only if $\varepsilon = h\nu$, where *h* is a constant that is independent of ν. At that point, Planck refrained from discussing the possible meaning of that constant, although he suspected that his 'quantum hypothesis' might have momentous (indeed, potentially devastating) consequences for theoretical physics. At any rate, his strategy in deriving the law had been inconsistent, for it had called upon both electromagnetism, in which the energy of an oscillator varies continuously, and a statistical treatment of that same energy, which supposes its discretisation...

Attempts to salvage the equipartition theorem and avoid a partitioning of energy into quanta continued well after 1900. All of them were unsuccessful, however, and Planck's radiation law remained without any serious competitor. Einstein, for one, did not regard the inconsistencies in Planck's derivation so much as flaws as hints that there was more to the microstructure of material bodies and to radiation than classical physics could account for. In one of the papers of the wondrous 1905 trilogy[2], Einstein challenged the very

---

[1] What is summarised here is, needless to say, the published end product of Planck's endeavour, which may not be faithful to the way he actually tackled the problem. Rosenfeld (1936) argues that Planck did actually work out an *ad hoc* expression for S_N that led to his earlier interpolation formula. Noticing that the argument of the logarithm was similar to the kind of combinatorial formula used in the calculation of the total number of possible complexions might *only then* have suggested to him a discretisation of the contributions of the resonators to the total energy.

[2] Significantly, Einstein called the content of that, rather than the celebrated 'relativity' paper, "sehr revolutionär".

foundations of our understanding of the nature of light and more generally, of electromagnetic radiation. His investigation focused on monochromatic radiation of a given frequency ν within an enclosure of fixed volume and in the range of frequency and temperature in which Wien's law was known to be valid. Requiring conservation of energy and that the entropy be maximal when cavity radiation is at equilibrium, Einstein showed that the change in entropy as the radiation, originally confined within a volume V_0, comes to occupy a volume V is

$$S - S_0 = k \log \left(\frac{V}{V_0} \right)^{\frac{E_\nu}{\beta \nu k}} \qquad (1.5)$$

(k is Boltzmann's constant and $E_\nu/V = u(\nu)d\nu$). The probability of finding *n* particles at any given time within a fraction V of the total volume V_0 is (assuming mutual independence) simply equal to $\left(\frac{V}{V_0} \right)^n$. Therefore, the change in Boltzmann entropy is

$$S - S_0 = k \log \left(\frac{V}{V_0} \right)^n , \qquad (1.6)$$

hence the identification $E_\nu = nk\beta\nu$. Comparison with Planck's law suggests setting the hitherto unspecified constant β equal to $\frac{h}{k}$. Einstein concludes that the spatial distribution of radiant energy is fundamentally discontinuous: monochromatic radiation behaves, at least in the small density domain where Wien's law holds, as if it were composed of a finite number of independent energy quanta, the magnitude of which is proportional to the frequency ν of the radiation.

Einstein's introduction of an elementary quantum of energy had the virtue of avoiding Planck's inconsistencies. On the other hand, if it was understood as promoting the idea of a granular nature of light, that conclusion then ran counter to conceptions that were by then well established. The wealth of evidence in favour of an undulatory nature of light did not imply that electromagnetic radiation should be devoid of any granular aspect. However, there were no grounds before Einstein for suspecting the existence, nor indeed even the possibility, of discontinuities in the distribution of radiant energy, never mind raise such discontinuities to the status of real corpuscles. Such corpuscles were not implied by the black-body law. In fact, Peter Debye (1910) would soon show that Planck's radiation law could be derived without making any reference to individual resonators. Given the assumption that energy itself is quantised according to the relation $E = h\nu$, Planck's law was

shown to follow as the result of optimising the number of distributions of energy quanta among the vibrations.

In any case, Einstein's successful explanation of the photoelectric effect was a strong argument in favour of the acceptance of light quanta (*Lichtquanten*). Ironically, an accidental manifestation of that effect had been reported, unwittingly, by Heinrich Hertz in the course of experiments (1887) that were hailed as the ultimate (!) confirmation of the validity of the continuist electromagnetic theory of light. From 1888, systematic studies of the 'anomalous' production of photoelectric current showed that the more electropositive the metal, the greater its responsiveness to light. A series of experiments conducted by Lenard between 1899 and 1902 showed that, for a given metal, the emission occurs only above a frequency threshold ν_0 and regardless of the intensity of the incident light. Moreover, the energy of the photoelectrons increases in proportion to the frequency ν of the incident radiation ($\nu>\nu_0$). Whereas none of those facts can be reconciled with Maxwell's theory, all of them find a straightforward explanation in terms of light quanta: the maximum kinetic energy of an extracted electron is equal to the difference between the quantity $h\nu$ supplied through its interaction with an individual 'quantum of light' and the amount of work needed for extracting an electron from the metal. Millikan's experiments (1916), which conclusively showed that the kinetic energy of the photoelectrons was, as predicted by Einstein, proportional to the frequency ν of the light absorbed by the metal, with a proportionality constant equal to Planck's, did much to change the status of the quantum idea from that of a rather baffling way of restoring empirical adequacy to that of a notion with a clearly accessible experimental correlate. The very idea of an interaction of electrons with individual quanta was suggestive of light being 'truly corpuscular', although the existence of the light corpuscles had long gone unnoticed. Later on, Arthur Compton's experiments[1] (1922) largely dispelled earlier doubts about Einstein's quanta. Since these conclusively appeared to be able to transfer momentum, they seemed to qualify as genuine 'particles', albeit particles of a very special kind[2]. Significantly, however, Bohr[3] acknowledged the 'existence' of light quanta[4] only in 1924.

---

[1] Compton interpreted his results by assuming that each X-ray radiation quantum of energy acted as a unit on a single electron, as if that quantum were concentrated in a single particle.

[2] If the velocity at which those quanta propagate is equal to the maximum admissible velocity imposed by the special theory of relativity, those particles are required to have zero rest mass.

[3] Bohr's reluctance to think of the notion of light quanta as anything but a heuristic device stemmed from his conviction that all interference phenomena could be interpreted only in terms of a wave-based picture of light – and after all didn't Planck's relation, connecting energy to frequency, appear to make an explicit reference to undulatory aspects?

[4] Einstein's quanta were only called *photons* after 1926 (the word was coined by G. N. Lewis).

If 'quantised' energy transfers were not confined to electromagnetism, then the quantum idea might well be relevant to reported 'anomalies' in the data gathered in a wider range of domains. On account of the shortcomings of the classical equipartition theorem, the molecular theory was certainly a candidate for revision. Einstein realised (1907) that the experimentally observed decrease of the specific heat of solids with temperature was indeed a case in point. Assuming that all the atoms in a solid oscillate with the same frequency ν, and neglecting interactions, Einstein used Planck's distribution to calculate the specific heat per mole. The resulting formula accounted for the 'anomalous' decrease while being, for high enough temperatures, in agreement with classical calculations based on the equipartition theorem (law of Dulong-Petit). This new instance of the effectiveness of the quantum idea made it clear that, whatever the consequences, its theoretical necessity had to be reckoned with. As Arnold Sommerfeld noted during the first Solvay Congress (1911), it would by then have been futile to keep thinking of the quantum as though it related to peculiarities of electromagnetic radiation. Physicists had, until 1905, incorrectly ascribed difficulties associated with the velocity of light to specific aspects of the electromagnetic field and its propagation. History would not repeat itself. Rather, one had to face the prospect of a complete refoundation of dynamics, with unforeseeable consequences in our understanding of molecular behaviour, radiation and atomic structure. That neglecting the constant h in Planck's radiation formula allowed the retrieval of that of Rayleigh-Jeans, or that the Dulong-Petit law followed from Einstein's at high temperatures, suggested that a prospective 'quantum theory' might call for basic amendments to the laws of classical (electro)dynamics. The very dimensions of h ('action') gave a faint hint at such future developments.

Another breakthrough was to concern no less than the stability of the atom. J.J.Thomson's model (1903), which had been dealt a fatal blow by scattering experiments with α particles, had been superseded by Rutherford's 'planetary' model (1911). Despite some attractive features, the new model was obviously unacceptable as it stood. One of its main drawbacks was that nothing in the laws of electrodynamics could possibly lead to the selection of definite periods and orbital distances. Thus, the orbits described by electrons around the nucleus were not constrained to have fixed radii, and the range of allowed periods was *a priori* continuous. But a far more serious problem was that the laws of mechanics and Maxwell's theory implied that electrons should irresistibly spiral down towards the nucleus and the atom collapse in a tiny fraction of a second (the rate at which the energy should be dissipated would be proportional to the square of the acceleration of the electron). The verdict was final: no classically conceived planetary atom could be stable.

Besides, patterns of spectral lines, characteristic of the elements involved, had been observed for a long time in the light emitted or absorbed by a variety of substances. By the end of the nineteenth century, arithmetical relationships between the positions i.e. wavelengths of lines had been worked out both by professionals and amateurs (notably by J.J. Balmer, a Swiss schoolteacher). Those regularities remained enigmatic, but there was little doubt they should relate to properties of material bodies on microscopic scales, and that their decryption could provide clues to atomic structure. At any rate, the very discreteness of those spectra flew in the face of classical physics, which would have had any spectrum consist of a fundamental vibration and higher harmonics[1]. Whatever could account for the spectroscopic data would therefore have to break away in some fundamental manner from classical treatments of matter and radiation.

Niels Bohr, who was freshly acquainted with Rutherford's model, was also aware of the role played by Planck's constant in constraining the magnitude of energy transfers. The quantum of action appeared to him as the best potential candidate for constraining energetic and orbital possibilities within each individual atom. His move was to hypothesise the existence, for each stable orbit, of a fixed definite ratio between the kinetic energy of an electron in orbit and its period of rotation. Supplementing with that hypothesis a planetary model of the mono-electronic hydrogen atom was tantamount to restricting permissible orbits to a discrete set. Stability was guaranteed by forbidding any radiation of energy as long as the electron remained in any one of the permitted (stationary) orbits. A transition between two energy levels corresponding to two distinct stationary orbits would be signalled by the emission of a single quantum of monochromatic radiation, according to the simple formula $E_m - E_n = h\nu$, where m labels a higher-energy orbit further from the nucleus than the 'lower' n-th (ν is, of course, the frequency of the emitted radiation). Why the orbits so selected were permitted while others were excluded was left unanswered, and no attempt was made to explain why or how a transition occurred and why it had to be accompanied by the emission of 'quantised' radiation.

Bohr's model of the hydrogen atom was a most unlikely hydrid, motion along orbits being governed by the classical laws of mechanics whilst the selection of permitted orbits and transitions between two such orbits would be regulated by principles that were entirely alien to classical electrodynamics. The gap could clearly not be bridged in the ideal case of circular orbits, for Bohr's hypothesis amounted to the assumption that the angular momentum of the electron round the nucleus would no longer be free to take any value in a

[1] Rutherford's model implied the production of continuous radiation by collapsing atoms.

continuous range, but had (?) to be an integral multiple of the ratio $\frac{h}{2\pi}$. Bohr's model was not explicitly based on the postulate of a quantised angular momentum. However, despite his reluctance to conjure up mechanical pictures, Bohr conceded that a 'symbolic' use of classical language could be useful in discussing the end product of what would otherwise be blind calculations. Provided that they were fully aware of the limitations of such concepts and terminology, physicists should feel free to refer to the amendments made to the planetary model in terms e.g. of a limit set to the possible range of values of the angular momentum of a bound electron.

On the basis of his oddly quantised planetary model, Bohr then gave three independent derivations of Balmer's empirical formula and predicted correct patterns of lines[1] for the 'Paschen series' (1908) and those to be observed later by Lyman (1914), Brackett (1922) and Pfund (1924). Further support was lent to his account of atomic transitions by the results of the experiment of Franck and Hertz (1914). Everyone, including Bohr, was well aware that his model was a crude step toward a more comprehensive and satisfactory account of the structure and stability of atoms. Further and lasting progress would come only through resolving or transcending the clash between the indubitable power of the quantum idea and the time-honoured laws of classical physics.

Designating the single angular coordinate (azimuthal angle) by φ, Bohr's implicit quantisation of the angular momentum can be expressed as the requirement that the angular component of momentum p_φ be such that $\oint p_\varphi d\varphi = n_\varphi h$ (integration being performed over a period). As it turns out, Planck's quantisation of the harmonic oscillator can be expressed in a similar way as $2\int_{q_-}^{q_+} pdq = nh$, with the momentum p satisfying $E = \frac{p^2}{2m} + \frac{1}{2}kq^2$ and $q_\pm = \pm\sqrt{\frac{2E}{k}}$ (the integral is then equal to $\frac{E}{\nu}$, where $\nu = \frac{1}{2\pi}\sqrt{\frac{k}{m}}$).

This suggested independently to Ishiwara, Sommerfeld and Wilson that discontinuous exchanges of energy might occur only between stationary configurations satisfying the quantum conditions $\oint p_k dq_k = n_k h$, where k ranges over all the degrees of freedom of the periodic system. Sommerfeld, in particular, applied those conditions to the treatment of the hydrogen atom. Realising that augmenting to two or three the number of degrees of freedom

[1] Bohr's predictions appeared to conflict with the observation by Pickering (1896), in the spectrum of the star ζ Puppis, of hydrogen lines that could not be accounted for by Balmer's formula. A brief controversy followed, which was settled when it was found (1914-1916) that the lines in question did not pertain to hydrogen but to helium (to which Bohr's model and Balmer's formula do not apply).

involved in the characterisation of the quantised 'planetary' motion did not give rise to an increase in the number of energy levels, he attempted a relativistic generalisation of Keplerian motion. As a result, there appeared a second-order 'correction' term, reflecting the imposition of the requirement of Lorentz invariance[1] (the term happened to be proportional to the square of a numerical constant $\alpha = \frac{2\pi e^2}{hc}$). Sommerfeld's extension of Bohr's original model was in good agreement with the observation of some of the finer structure of atomic spectra (as measured, e.g. by Paschen circa 1916), hence the name *fine structure constant* usually given to α.

Paul Ehrenfest (1906) was probably the first to realise why Wien's displacement law[2] survives the Planck-Einstein 'quantum' treatment of black-body radiation: it is essentially because it is (tacitly) assumed that the radiation enclosure undergoes infinitely slow compression. The motion of a system acted upon by external forces is said to be 'adiabatic' if those forces do not directly affect the coordinates of the system when relevant parameters vary slowly. In the case of the harmonic oscillator, looking for invariants boils down to finding some function of the parameters and constants of the motion that should be invariant under the transformation. That function should, in particular, remain equal to the initial value of the energy-to-frequency ratio of the oscillator, i.e. to *nh* in accordance with Planck's quantum hypothesis. Ehrenfest showed (1916) that Sommerfeld's phase integral conditions, as applied to the hydrogen atom, were just an expression of adiabatic invariance[3]. Why circularity (Bohr) or ellipticity (Sommerfeld) of the orbits makes no difference to the spectrum also found an explanation: the elliptic and the circular case are related by an adiabatic transformation over the same time period. Given the square law form of the Coulomb potential, the total energy must be the same in both cases (virial theorem).

---

[1] A special relativistic treatment involving the essentially 'prerelativistic' Coulomb form of the potential is as highly suspicious as Planck's incongruous blend of continuum physics and combinatorial methods. Darwin's later (1920) treatment restored consistency by substituting a retarded potential for the Coulomb term.

[2] The ratio of each proper vibration of frequency ν to that frequency is a unique function of ν and temperature.

[3] The ratio $\frac{\varepsilon_\nu}{\nu}$, equal to *nh* according to Planck, is an adiabatic invariant, and so is the ratio $\frac{\nu}{T}$. In the simple case of Bohr's circular orbits, Bohr's condition $m\omega r^2 = \frac{h}{2\pi}$, where $m\omega r^2$ is the angular momentum (m: mass of the electron, $\omega = 2\pi\nu$: angular velocity, r: distance from nucleus) can be written $\frac{2\bar{K}}{\nu} = h$ in terms of the mean kinetic energy $\bar{K} = \frac{1}{2} m\omega^2 r^2$, and $\frac{2\bar{K}}{\nu}$ is precisely the adiabatic invariant that is equal to Sommerfeld's phase integral.

In contrast to Kepler's planetary problem, the quantum conditions, hence the quantised orbits, are sensitive to the choice of coordinates. How could a given choice be justified, and was there one that was most adequate? The question was answered by the astrophysicist Karl Schwarzschild, who was well versed in the technicalities of celestial mechanics (another important contribution to the same problem was made by Paul Epstein, who had been a student of Sommerfeld). Roughly, the idea was to take advantage of a 'global method' developed in the previous century for computing the frequencies of periodic motions, whilst bypassing the need to completely solve the equations of motion. The method, introduced by Delauney in his *Théorie du mouvement de la lune* (1860-1864) and used extensively by Poincaré, was little known outside the circle of expert astronomers (unsurprisingly, since the question of the exact nature of the motion was left aside). This computational approach involves the use of so-called *action and angle* variables, whose defining relationships bear a striking formal resemblance to key quantities in the Bohr-Sommerfeld account of atomic structure. Schwarzschild realised that exporting those techniques from their original domain could prove very useful to the development of atomic physics. The success of Sommerfeld's calculations owed much to the possibility of separating variables in the Hamilton-Jacobi equation. Practitioners of celestial mechanics had known for at least three decades that systems with a finite number N of degrees of freedom, for which the Hamilton-Jacobi equation could be integrated using separated variables, are multiply periodic, in the sense that the generalised coordinates can be expressed as a N-fold infinite Fourier series. This, naturally, did not imply that the motion itself had to be periodic (it is clearly so only if frequencies in the expansion are commensurable). Action-angle variables can be substituted for every (q_k,p_k) pair, letting $J_k = \oint p_k dq_k$ (integration being performed over a complete rotation or libration[1] of the coordinate q_k) for the 'action' and $w_k = \frac{\partial W}{\partial J_k}$ for the 'angle' variable, where W is the generating function for the $(q,p) \rightarrow (w,J)$ transformation[2]. The similarity between J_k and Sommerfeld's phase integral is evident. Epstein argued that Sommerfeld's generalised quantum conditions *had* to be formulated using coordinates that would allow the Hamilton-Jacobi equation to be separated, i.e. in terms of action and angle variables. What conditions should then be satisfied by dynamical systems with time-independent Hamiltonians for action and angle variables appropriate for the formulation of the quantum conditions to be introduced? An answer, found

[1] Oscillation between fixed limits (the initial position lies between two zeroes of the kinetic energy).

[2] W is Hamilton's characteristic function if the system is conservative (see Goldstein 1980, p.460).

by Burgers (1916-1917), was that the action variables should be adiabatic invariants. If anything, those technical developments suggested that progress toward the elaboration of a 'quantum-mechanical' framework might require probing the mathematical depths of analytical mechanics.

Another important heuristic guide was provided by the so-called 'correspondence principle' – if, indeed, there is any single such principle[1]. Classical expressions were expected to hold when a dimensional analysis showed the order of magnitude of typical 'actions' to far exceed the value in the same units of Planck's constant ($h \approx 6.62.10^{-34}$ Joule-second). Practically, *h* is treated as an adjustable parameter (which it is of course not) and classical expressions are 'recovered' as $h \to 0$. Besides, asymptotic agreement between the frequency spectrum and the classical description of the motion is expected in those regions further away from the nucleus where the difference in energy between any two stationary orbits becomes negligible, i.e. when the 'principal quantum number' *n*, which labels energy levels, goes to infinity[2]. Following Bohr, who had long given up any hope or intention of relating discrete radiation transfers to the harmonic components involved in a classical description of orbital motion, many came to regard asymptotic correspondence as the only available handle on the problem of deriving correct intensities and polarisations of spectral lines.

In a 1916 paper, Einstein made another significant contribution to the treatment of 'quantised' electromagnetic radiation. For the very first time, Einstein assigns *a priori* probabilities to transitions, whether induced or not, from a higher energy level (E_m) to a lower one (E_n), and those (induced) from a lower to a higher level. Starting from the assumption that the canonical distribution is preserved under conditions of equilibrium, Einstein derives Bohr's relation $E_m - E_n = h\nu$ and finds how the probability p_{mn} of non-induced emission relates to that of induced emission for the same two levels: $p_{mn} = \frac{8\pi h\nu^3}{c^3} p_{mn}^{(induced)}$. The cubic dependence on frequency is characteristic of Planck's radiation law. There is little doubt that Einstein regarded those probabilities as reflections of our limited knowledge of structural rearrangements that really take place in the atom. His later overt hostility to 'quantum probabilism' also dispels any suspicion that he could have ever conceived of individual transitions as intrinsically undetermined. Bohr, on the other hand, was already convinced of the impossibility of furthering our understanding of atomic structure beyond a kind of theoretical guesswork, where classical physics was to be used as a merely heuristic but

---

[1] See next footnote.

[2] The question of consistency between the two 'limits' $h \to 0$ and $n \to \infty$, and the associated forms of correspondence, is not trivial ; see Liboff 1984.

indispensable guide. Given his personal outlook, Bohr readily interpreted as 'spontaneous[1]', in the strong sense of acausal, those transitions Einstein merely assumed were 'not induced' by external influences. In view of the considerable ideological influence of Bohr on the development and final acceptance of quantum mechanics, what is implied in his statement of a basic and irreducible indeterminacy at the heart of quantum physics cannot be underestimated. From there indeed, it was but a relatively short step to think of the subject matter of quantum theory in terms of 'virtual' processes or 'potential' occurrences, brushing aside the possibility that the interactions and behaviour of microphysical objects remain amenable to a fully mechanical treatment – if no longer along classical lines. In a letter to Høffding (1922), Bohr confided his assessment of the situation:

> ...we encounter difficulties which lie so deep that we do not have any idea of the way of their solution; it is my personal opinion that these difficulties are of such a nature that they hardly allow us to hope that we shall be able, inside the world of the atoms, to carry through a description in space and time of the kind that corresponds to our ordinary images.

Progress toward a comprehensive and adequate quantum theory, which could encompass both radiation and atomic structure, was still wanting by 1924. Tension between discrete and continuous aspects raised the question of 'gaps' that might exist in energy and momentum transfers and might threaten their exact conservation. Early in 1924, John C. Slater expressed his view that, if a connection was to be made between undulatory features and light quanta, then in order to break with classical ideas of causation and circumvent difficulties with conservation laws, that connection would have to be statistical. Bohr, who was apparently keen to renounce any causal description of atomic transitions, was responsive to Slater's suggestions. Slater, Bohr and Hendrik Kramers developed their programme in a paper published in the same year. Their basic idea was to associate with every stationary atomic configuration a 'virtual' radiation field. What made that field *virtual* was its being composed of as many monochromatic (spherical) waves as there were *possible* transitions to lower energy levels in the atom. In other words, rather than be defined in terms of whatever is thought to be the case, that field would be defined in terms of what *could* be the case. Since there were as many virtual oscillators as spectral lines and there was no obvious correlation between orbital motion (or whatever was left of it) and the virtual oscillators, which were as numerous as spectral lines, transitions in distant atoms could not be causally connected: the absorption of a quantum by one atom A_1 could be induced by the virtual radiation field associated with another atom A_2 without implying that A_2 should undergo any 'downward' transition (with emission of one quantum).

[1] The word, though not Einstein's, has stuck, as sampling textbooks will readily confirm.

With Kramers and Slater, Bohr had ventured further along the track opened by his personal interpretation of Einstein's probabilities. In accordance with the contention that no mechanical picture of transitions could be given, the scheme would only yield a probability for a given transition to occur during a specified time interval. That probability appeared to be nothing but the expression of a ghostly conspiracy of possibilities. The acausal features of the theory met the disapproval of those who, like Einstein, were not ready to give up strict causality, and the scepticism of most. Pauli dismissed the enterprise as both artificial and misguided. Its experimental disproof by Bothe and Geiger in 1925 was certainly welcome with some relief.
Kramers went on developing a theory of dispersion (1924) where the faded echo of the virtual oscillators could still faintly be heard. His mathematically detailed study of dispersive behaviour led him to a tabulation of transition frequencies that we can now regard as the earliest step towards a matrix formulation of quantum mechanics. At about the same time Max Born, in a paper titled '*Über Quantenmechanik*' (1924), showed that, as a result of orbital quantisation and since the classical fundamental frequency is equal to the derivative of the Hamiltonian with respect to the action variable, classical formulas could be translated, as it were, into their quantum analogues by means of a simple rule, amounting to the replacement of a differential by an appropriate finite difference.

A newcomer in the quantum arena, Werner Heisenberg had been a keen listener when Bohr was lecturing at Göttingen on the latest developments in atomic physics. Bohr's insistence on 'fitting' the needed concepts to the available experimental evidence deeply resonated with Heisenberg's inclinations. Another influence on his outlook was Einstein's discussion of a proper usage of 'clocks and rods', as a result of which age-old tenets had been overthrown. Heisenberg found in Einstein's groundbreaking analyses an incentive to concentrate his own efforts only on quantities that were experimentally accessible. His radical answer to the current muddle was to reject any attempt at picturing atoms and their transformations, and to focus instead on quantities of direct relevance to the calculation of transitions between specified initial and final atomic configurations. Asking, and trying to answer, idle questions about 'interphenomena' could, in his view, only delay progress toward the elaboration of a valid framework for atomic physics.
Any attempts to base a viable account of atomic phenomenology on a 'microscopic kinematics', in which the electrons would be characterised at any time by well-defined values of position and momentum, had so far been unsuccessful. Heisenberg contended that optical quantities like frequency or dipole amplitude, which could be directly related to the observational data, should be taken as a springboard for deriving a consistent and effective

theoretical scheme. If anything was to remain, if only in a 'symbolic' manner, of the planetary model, optical frequencies of spectral lines might have been expected to coincide with definite aspects of the periodicity of orbital motion. Harmonic (Fourier) analysis of that motion into orbital frequencies did not, however, match the spectroscopic data. Heisenberg hypothesised that 'quantum' transition frequencies ν_{mn} should somehow 'correspond' to classical Fourier frequencies $\nu(m,n)$, as these occur in the Fourier expansion $\xi_m(t) = \sum_n x(m,n)e^{2\pi i\nu(m,n)t}$ of a time-dependent classical magnitude $\xi_m(t)$. More boldly still, he assumed that a similar correspondence should also hold between each Fourier amplitude $x(m,n)$ and some quantum-theoretic quantity $x_{m,n}$, which was as yet purely hypothetical. Even if some physical meaning could be ascribed to each $x_{m,n}$ separately, it was then far from clear whether one could make any sense at all of the 'quantum-theoretical' analogue $\sum_n x_{m,n}e^{2\pi i\nu_{m,n}t}$ of the Fourier series of $\xi_m(t)$. Heisenberg's answer was to assume that, even if no straightforward correspondence could be thought of, the 'collective' (*Gesamtheit*) of individual terms $x_{m,n}e^{2\pi i\nu_{m,n}t}$ forming the above sum could at least be chosen as a valid quantum-theoretical 'representative' of the classical quantity $\xi_m(t)$. If it meant anything, this peculiar correspondence should satisfy consistency constraints. Thus, if a given collective was to be a valid representative of some quantity ξ_m, the square ξ_m^2 of that quantity should itself admit a representative *Gesamtheit* whose frequencies and 'amplitudes' properly meshed with those in the 'ξ_m-collective'. This, Heisenberg concluded as a result of some algebraic manipulations, would be guaranteed provided that the $x^{(2)}_{m,n}$ coefficients in the 'ξ_m^2-Gesamtheit' satisfied the composition rule $x^{(2)}_{m,m-n} = \sum_k x_{m,m-k}x_{m-k,m-n}$. The representative of a product $\xi_m\zeta_n$ of two classical quantities ξ_m and ζ_m should satisfy a generalisation of this rule. However, the resulting coefficients would then in general differ from those associated with the product $\zeta_m\xi_m$ – in other words, the composition rule was *not* commutative.

The dynamical part of Heisenberg's framework had yet to be worked out. If their classical analogues are solutions of Newton's equation of motion, subject to Sommerfeld's quantisation condition, Heisenberg's $x_{m,n}$ would have to satisfy appropriate relations. After deriving them[1], Heisenberg used those relations to get the expressions for the energy levels of 'quantum' harmonic

[1] Those relations comply with Born's 'translation rule' which, it will be recalled, substitutes a finite difference for a differential. Whilst working under Born's supervision, Heisenberg had been acquainted to that particular rule and had made use of it in his contribution to the dispersion theory initiated by Kramers.

and anharmonic oscillators. Unlike calculations performed with the older methods, these expressions implied the existence of a 'residual', so-called zero-point energy. The prediction was in full agreement with experimental results (these had been actually obtained a few months before the submission of Heisenberg's paper, at which time he was not aware of them).

Heisenberg's approach and procedures were bold and controversial. Nevertheless, their importance was recognised almost immediately by Bohr. What is perhaps more surprising is Bohr's praising them as a "precise formulation of the tendencies embodied in the correspondence principle", despite Heisenberg's mistrust of classical representations and his postulation of abstract 'sets' or collectives of time-dependent – and complex – numbers. Born, who qualified like none other for assessing the technical merits of Heisenberg's paper, soon realised that Heisenberg's odd composition rule for amplitudes was nothing but the simple rule of matrix multiplication. That this could have escaped the attention of Heisenberg himself is understandable, for matrices were not yet part of the usual toolbox of the physicist (Heisenberg knew nothing about them). It so happened, however, that Born was familiar enough with matrices to have used them in his own work on the lattice structure of crystals (1921). He also realised that further progress would call for expertise in handling those unfamiliar tools. A chance meeting on a train made Born lay hand on the right man at the right time: Pascual Jordan not only had a solid experience in matrix calculus[1]; he was also eager to assist him in his research programme. Born and Jordan almost immediately teamed up and, just two months after Heisenberg's work had been published, produced a rigorous account of what they referred to as *Quantenmechanik*. Heisenberg's collectives had been duly replaced with matrices **q** and **p**, thought of as 'representatives' of position (spatial coordinate) and momentum (classically conjugate to that coordinate) respectively.

A matrix analogue of the classical Lagrangian being tentatively written $\mathbf{L} = \mathbf{p}\dot{\mathbf{q}} - \mathbf{H}$, where $\dot{\mathbf{q}}$ is the matrix whose entries are the time derivatives of the entries of **q** and **H** is a matrix 'Hamiltonian' function of **p** and **q**, extremising the trace of **L** (a 'quantum-mechanical' adaptation of the usual procedure of extremising the Lagrangian action in classical mechanics) leads to a 'matrix form' of Hamilton's canonical equations:

$$\dot{\mathbf{q}} = \frac{\partial \mathbf{H}}{\partial \mathbf{p}} \quad , \quad \dot{\mathbf{p}} = -\frac{\partial \mathbf{H}}{\partial \mathbf{q}} . \qquad (1.7)$$

[1] Jordan had assisted Courant in the preparation of the monumental *Methoden der mathematischen Physik*, which were based on Hilbert's lectures. Algebraic concepts and methods, expounded in the first volume (1924), would soon prove essential to the mathematical foundations of quantum mechanics.

Following Heisenberg, but in full awareness of the matrix nature of his 'collectives', Born and Jordan prove that those matrices must satisfy the simple 'canonical' commutation relation[1]

$$\mathbf{pq - qp} = \frac{h}{2\pi i}. \qquad (1.8)$$

By the end of 1925, the *'Dreimännerarbeit'* of Heisenberg, Born and Jordan led to the first comprehensive framework – known for a while as *matrix mechanics* – that could rightly be called a quantum theory. Bohr's frequency condition and all other experimentally vindicated features of earlier treatments of atomic structure or radiation could be derived from the new scheme. The methodology, however, stood in total contrast with that of classical problem-solving, for it required finding two time-independent (Hermitian) matrices that satisfy the commutation relation, and such that the Hamiltonian matrix for the system is diagonal. In formal analogy to Hamiltonian mechanics, canonical transformations could be defined with the purpose of facilitating the resolution of the 'equations of motion' – if the new 'quantum-mechanical' equations could at all be regarded as 'describing' any motion... The powerful 'Hilbert-Courant' algebraic methods found successful application in the calculation of matrix elements (e.g. diagonalisation) – to the extent, at least, that those methods were actually applicable to infinite, unbounded matrices (which was by then not established).

There again, much of the mathematics used in the new framework had found their first application in celestial mechanics[2]. Even if the approach initiated by Heisenberg was based upon the rejection of any representation of electrons in orbital motion, some of the mathematical techniques which had been developed to deal with genuine planetary orbits were none the less taken advantage of. Besides, if mutual stimulation between mathematics and physics had not waited for the quantum, it also seemed to have entered a new era, and it was clear enough that matrices were just the tip of an iceberg. Drastic departures from classical concepts and procedures had been made in the course of working out the new formalism, e.g. through the early imposition of Sommerfeld's action quantisation or the later substitution of a difference for a differential. Whether it was appropriate to conceive of the new framework as a new kind of *mechanics* could therefore be questioned. Besides, the physical necessity of Planck's constant remained an enigma.

---

[1] Heisenberg had actually left momentum out of his *Gesamtheit* representation. If he had not omitted it, it is likely he would have been the first to get to the canonical commutation relation (1.8), although without recognising its matrix form.

[2] Perturbation methods were initially developed to meet the needs of the astronomers, e.g. for resolving complex planetary motions into stable orbits. The prediction of the existence of Neptune, by Leverrier and by Adams, is a famous instance of their successful use.

In spite of its success, the new framework had inherent limitations: it had been constructed specifically for tackling problems in atomic physics where periodicity was essential – even if any kind of orbital imagery was definitely out. Ironically enough, the sophisticated machinery of matrix mechanics was ill suited for dealing with the simplest kind of aperiodic motion encountered in classical mechanics: uniform rectilinear motion. Born took advantage of an invitation to the MIT to discuss those matters with Norbert Wiener, whom he had met at Göttingen the previous year. Wiener's proficiency in Fourier analysis was an asset the physicist and mathematician made full use of, developing the first account of quantum theory in terms of linear operators. The Born-Jordan commutation relation (1.8) and quantum analogues of the classical canonical equations were reformulated as relations between operators with Hermitian matrix representatives. After introducing an energy differential operator $\mathbf{H} = \frac{h}{2\pi i}\frac{d}{dt}$, Born and Wiener applied their operator-based formalism to such basic periodical problems as the harmonic oscillator, but also to aperiodic situations like that of rectilinear uniform motion. An operatorial formulation was more manageable than the more cumbersome matrix-based scheme, and certainly more promising as a first step towards understanding the hows and whys of the effectiveness of the new theory. Oddly enough, Born and Wiener failed to notice that the canonical commutation relation was precisely of such a form as to suggest writing **p**, just like **H**, as a differential operator: $\mathbf{p} = \frac{h}{2\pi i}\frac{\partial}{\partial q}$. Had they made that one step, they could have easily worked out, from very different premises, the kind of differential equation that Schrödinger would come up with just a few weeks later.

Unaware of those developments, Paul Adrien Maurice Dirac, then in Cambridge, took a personal look at Heisenberg's multiplication rule (which he also failed to identify, however, as an instance of matrix multiplication). Treating any quantum-mechanical variable x as a classical function of action and angle variables, and expanding it as a Fourier series under the assumption that the action is quantised in multiples of Planck's constant, he found that the difference xy – yx was formally analogous to the Poisson bracket[1] of x and y in classical mechanics, and that (1.8) immediately follows from the Poisson brackets for position and momentum variables. It will be remembered that Born and Jordan had first obtained 'quantum' analogues of Hamilton's equations from the application of a variational principle to a matrix

[1] The classical Poisson bracket of two classical magnitudes x and y is

$$\{x,y\} = \frac{\partial x}{\partial \omega}\frac{\partial y}{\partial J} - \frac{\partial y}{\partial \omega}\frac{\partial x}{\partial J},$$

where ω is the angle and J the action variable.

version of the classical Lagrangian, and then derived (1.8) from there. Since Poisson brackets are implicit features of Hamilton's formulation of mechanics, it is little wonder that another (operator-based) route to the same 'quantum conditions' (as Dirac would later[1] call them) should give rise to a parallel between Poisson brackets and quantum-mechanical commutators.

Finally, Dirac obtained the expression of the dynamical equation a function **F** of the quantum-mechanical canonical variables **p** and **q** should satisfy:

$$\frac{d}{dt}\mathbf{F} = \frac{2\pi}{ih}(\mathbf{FH} - \mathbf{HF}), \qquad (1.9)$$

where **H** is the quantum-mechanical Hamiltonian (operator), which in analogy with classical physics represents the energy of the system. This is the equation that regulates the time evolution of a dynamical variable **F** in the so-called 'Heisenberg picture'.

Although he did not explicitly mention operators as the proper mathematical representatives of quantum-mechanical quantities, there was no doubt in Dirac's mind those 'q-numbers' (as he referred to them in a subsequent paper) *should* be substituted for the more familiar commuting 'c-numbers' of classical physics[2]. However, the need for such substitution did not weaken his conviction that any deep understanding of the formalism should make obligatory reference to the Hamiltonian formulation of classical mechanics.

Quite independently, another trend of thought had crystallised around Einstein's notion of light quanta. Louis de Broglie, a doctoral student at the Sorbonne (1923), looked for a way to reconcile a corpuscular view with periodic features that were apparently part and parcel of the quantum hypothesis. He imagined that a moving particle of internal energy m_0c^2 would be the seat of vibrations with frequency ν_0, such that $m_0c^2 = h\nu_0$. The moving object would also be accompanied by a (fictitious) wave – "*une onde fictive associée au mouvement du mobile*" – that would keep in phase with the internal periodicity of the body. The latter assumption was introduced in order to eliminate a possible discrepancy between ν_0 and the frequency observed by a stationary observer. The wavelength of the wave associated with a particle of mass m and total energy E, subject to a potential V, would be $\lambda = \frac{h}{p} = \frac{h}{\sqrt{2m(E-V)}}$, where p was the particle's momentum. De Broglie applied his proposal to the motion of an electron on a closed trajectory, e.g. an atomic orbit. If the accompanying wave described the same orbit, phase agreement between the wave and the intrinsic vibratory process within the electron led to a stability condition that just happened to be Sommerfeld's.

[1] Dirac 1958.
[2] Darrigol 1992.

The group velocity of the phase waves would then be equal to the actual velocity of the particle on its trajectory.

The idea that a particle in motion would, at each point of its trajectory, follow the normal to a surface of equal phase at that point was familiar to Louis de Broglie. This was in fact a key aspect of a most intriguing analogy between optics and mechanics, suggested by William Rowan Hamilton in the 1830s. Just as Lagrange had regarded mechanics as a branch of mathematics, Hamilton was not so much concerned with the nature of light or motion as with developing a 'harmonious' formalism that would bring together, for the satisfaction of intellect, two seemingly disconnected domains of experience. It so happens that mechanics and geometrical optics were both regulated by variational principles: mechanics by a principle of stationary action, and optics by Fermat's. In mechanics, the (configuration space) path that connects two fixed configurations $\gamma(t_1)$ and $\gamma(t_2)$ of a system at times t_1 and t_2 satisfies Newton's laws of motion just in case that path makes the time integral of the Lagrangian L stationary with respect to (δ) variations of the path that vanish at the end points: $\delta\int_{t_1}^{t_2} L dt = 0$. Considering only systems for which the total energy E is a constant and the potential V does not depend on time, letting $L = T - V$ and $T+V = \frac{p^2}{2m} + V = E$, the stationarity condition reduces to $\delta\int_{t_1}^{t_2} T dt = 0$ or $\delta\int_{M_1}^{M_2}\sqrt{2m(E-V)}ds = 0$, where ds is an infinitesimal element of the mechanical path. Fermat's principle, which states that the actual optical path extremises the time taken by light to propagate between the end points, is formally expressed in a similar way as[1] $\delta\int\frac{ds}{v} = 0$.

Consider a single point-like mechanical system (a 'particle') of mass m, its position being labelled using rectangular (cartesian) coordinates (x,y,z). Hamilton's principal (S) and characteristic (W) functions are related by $S = W - Et$. Any surface of constant S must therefore coincide, at any given time, with some surface of constant W. As time goes by, every surface of constant S can be thought of as displaced to another locus in configuration space. The surface of action propagates similarly to a wave front, undergoing deformations in accordance with local variations of the potential. The velocity *u* with which the wave front propagates – again, in the relevant configuration space, which only accidentally coincides with three-dimensional physical

[1] Here, ds is an infinitesimal element of the optical path, $v = \frac{c}{n}$ is the phase velocity and n the refractive index of the medium.

space in the single particle case – can be defined as the ratio of the infinitesimal distance ds normal to the surface, to the infinitesimal time required to cover that distance: $u = \frac{ds}{dt}$. The infinitesimal variation of W during an infinitesimal time interval dt is dW = Edt. W is a scalar function, hence $dW = \left|\vec{\nabla}W\right| ds$ and $(\vec{\nabla}W)^2$ = 2m(E-V) is the stationary Hamilton-Jacobi equation. The velocity of the action wave is therefore $u = \frac{E}{\left|\vec{\nabla}W\right|} = \frac{E}{\sqrt{2m(E-V)}}$, hence a function of the potential[1]. Since the components of the momentum p of the one-particle system are the partial derivatives of W relative to the spatial coordinates, i.e. $p = \vec{\nabla}W$, the momentum vector is orthogonal to surfaces of equal action in configuration space[2]. At any time during the motion of the particle, the direction of its configuration path will be orthogonal to a constant action surface. Between two configuration end points M_1 and M_2, the motion of action wave fronts must satisfy the variational condition $\delta \int_{M_1}^{M_2} \frac{ds}{u} = 0$, in close analogy to Fermat's principle. As expected however, there are fundamental discrepancies between the optical and the mechanical accounts. Thus, in geometrical optics, light propagates more rapidly where the corresponding wave fronts are far apart, whereas action wave fronts propagate slowly while the particle moves fast i.e. where surfaces are tightly packed (u is inversely proportional to the speed of the particle). Hence, if a wave motion can be said to 'accompany' the change of configuration with time of a (N-particle) system (in the relevant 3N-dimensional space), the two 'associated' motions do not keep step. The association of a wave with a particle, or with a multiparticle system, is fundamentally formal – particles do not 'ride' action waves.

If Hamilton's proposal failed to capture the interest of his contemporaries, it is mainly because geometrical optics is, after all, only an approximation to undulatory optics. A parallel between its variational expression (Fermat's principle) and a variational formulation of mechanics is intriguing – is it really significant? Decades later, the appearance of the quantum raised another, deeper issue: could classical mechanics be to 'wavelike' motion what geometrical optics is to undulatory optics, i.e. a short-wavelength limit?

---

[1] u = E/p in the single-particle case, since the momentum p is such that $p^2 = 2m(E - V)$, where $E - V$ is the kinetic energy.

[2] Why it has to be so can easily be figured out: if the particle path 'crosses' two neighbouring surfaces of constant action S_1 and S_2 at M_1 and M_2 respectively, the orthogonality of momentum to both surfaces will imply *less* action than there would correspond, say, to a path connecting M_1 to *any other* point M'_2 on S_2 (the difference in action is precisely given by $pds = \sqrt{2m(E - V)}ds$, where ds is the infinitesimal distance between the two surfaces and p the absolute value of the momentum).

Perhaps undulatory aspects, which are negligible or averaged out in all macroscopic situations, can no longer be overlooked as physicists investigate finer levels of material structure and their coupling to radiation[1]. De Broglie's speculative ideas did not create much of a stir in the scientific community. Nevertheless, Einstein's acute vision perceived in them the seed of a conceptually more rewarding alternative to the dazzling computational techniques of matrix mechanics. What de Broglie's work seemed to suggest was that what was perhaps needed was the derivation of an 'action wave' equation, which would be to the classical equations of mechanics what the wave equation of optics is to the basic 'eikonal' equation of geometrical optics. Schrödinger's attention had been drawn to de Broglie's thesis whilst reading one of Einstein's papers[2]. He realised that de Broglie's derivation of Sommerfeld quantisation as a constraint on the number *n* of wavelengths that 'fit' within an atomic orbit ($n = \oint \frac{1}{\lambda} dq$, with $\lambda = \frac{h}{p}$) could be treated as an eigenvalue problem. His first application of that treatment to the bound electron was quite unsuccessful, but he could not by then have been aware that the discrepancy was due to his overlooking electron spin. He concluded instead that his special-relativistic treatment of the problem was flawed. Invited by Debye, his predecessor at the University of Zurich, to organise a colloquium on de Broglie's ideas, Schrödinger re-examined his initial calculations and found that they agreed with experimental data in the Galilean-relativistic approximation. In a seminal paper published in 1926, Schrödinger elaborates at length on Hamilton's analogy, carrying it further that de Louis de Broglie had done, or was indeed ever willing to do.

In the approximation of geometrical optics, the wavelength of the propagating light is assumed to be small compared with distances over which the refractrive index *n* of the medium varies. When the index is constant, the equation for wave propagation admits a plane wave solution with phase[3] $k.r - \omega t$. If the index varies, the wave will undergo bending and distortion and the wave equation will no longer admit a plane wave solution. However, under

---

[1]Clues to undulatory features could actually have been spotted in the available data as early as 1924, i.e. before they were clearly revealed in the results of scattering experiments performed with electrons by Davisson and Germer (1925) An interpretation of the latter in the light of de Broglie's thesis (or, by then, of Schrödinger's version of quantum mechanics) was actually delayed until 1927. It is also somewhat ironical that the final confirmation of diffraction effects involving electrons should have come from experiments performed by J.J. Thomson's son: three decades after the father had 'conclusively shown' that electrons 'were' particles, George P. Thomson showed that they could 'behave like waves'...

[2] This is the 1925 paper in which Einstein elaborates on S.N. Bose's original treatment of an assembly of light quanta, and derives Plancks' radiation law on purely 'statistical' grounds. Einstein briefly mentions de Broglie's ideas which, in his opinion, "involve more than merely an analogy".

[3] The wave number k, wavelength λ and angular frequency ω are linked by $k = n\omega/c = 2\pi/\lambda$, where c is the speed of light in vacuo.

a very gradual change of *n* the solution will retain a similar form as the plane wave. In particular, the total phase can be written $k_0(L(r) - ct)$, where L is a real function of position that is usually called the *eikonal* or *optical path length* (k_0 is the wave number in the vacuum). Imposing such a form to the solution of the wave equation and retaining only those terms with the highest power of k_0 leads to the eikonal equation of geometrical optics: $(\vec{\nabla} L)^2 = n^2$. Hamilton's analogy originates in the observation that this and the Hamilton-Jacobi equation $(\vec{\nabla} W)^2 = 2m(E-V)$ for the motion of a single particle in a potential V have the same basic form. On the mechanical side, the characteristic function plays a similar role as the eikonal does in geometrical optics. Action then plays the role of phase, the momentum that of the wave vector and the mechanical energy that of the frequency. Comparing the total phase of a light wave with the relation S = W – Et, taking into account Planck's relation E = hν, suggests writing a separable solution for quantum action waves in the form $\Psi(q,t) = \psi(q)e^{2\pi i(\frac{E}{h})t}$. 'Plugging' it in the wave equation $\nabla^2\Psi + nk_0{}^2\Psi = 0$ then leads to[1]

$$\nabla^2\psi + \frac{2m}{\hbar^2}\psi = 0. \qquad (1.10)$$

In fact, this equation appears, prior to its formal derivation, in the first part of Schrödinger's paper, devoted to the hydrogen atom. Schrödinger substitutes for the principal function S a term that is proportional to the logarithm of some function ψ, whose separable form lends itself to analytical treatment. The resulting equation, of the Hamilton-Jacobi type, is written as a quadratic form in ψ and its derivatives, where ψ is required to be real, twice continuously differentiable, and extremal when the quadratic form is integrated over the whole of the *q-Raum* (configuration space). The Euler-Lagrange equation associated with the variational problem is then the sought wave equation. For negative values of the energy E, the equation has solutions only provided energy is quantised i.e. takes discretely distributed values $E = -\frac{me^4}{2K^2}n^2$ (n ≥1 being an integer). The result is in accordance with the Bohr energy spectrum provided K is set to the value $\hbar$. A discrete energy spectrum thus emerges as a result of requiring that ψ functions, solutions of (1.10), have suitable mathematical properties. In other words, Schrödinger substitutes formal requirements for the (perhaps no more) enigmatic postulation of a quantisation of energy or angular momentum. Applied to the harmonic or

[1] The convenient symbol $\hbar = \frac{h}{2\pi}$ was introduced after 1926.

Planck's oscillator, the theory yields the same results (e.g. zero-point energy) as Heisenberg's procedure, although the route is quite different.
The third installment of Schrödinger's *Quantisierung as Eigenwertproblem*[1] paper is devoted to the development of an appropriately 'wave-mechnical' perturbation theory, which is successfully applied to the Stark effect. It is there that the general equation

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi + V\Psi \qquad (1.11)$$

is introduced, the solution of which is a time-dependent function $\Psi(q,t)$. This equation has the typical form of the equation for a diffusion process, albeit with an imaginary coefficient. This similarity eventually leads Schrödinger to relax the requirement that the Ψ 'wave function' should be real-valued. Thus, not only are the solutions of Schrödinger's equation confined within the abstract arena of configuration space, those solutions are now generally complex-valued.
De Broglie, unlike Born and Wiener (in a different context), had not just failed to realise he was but a short step away from writing down the equation for 'quantum wave' propagation. But there were waves and 'waves': if it had to 'evolve' in configuration space, this in his view would have amounted to nothing but denying the wave its 'physicality':

> L'idée de M. Schrödinger de définir l'onde Ψ d'un système dans l'espace de configuration m'avait au début beaucoup scandalisé parce que, l'espace de configuration étant purement fictif, cette conception enlève à l'onde toute réalité physique : pour moi, l'onde de la mécanique ondulatoire devait évoluer dans l'espace physique à trois dimensions[2].

If Schrödinger's breakthrough was not to de Broglie's taste others, including Einstein, Planck and Sommerfeld were quite enthusiastic. At last, it looked as if all the puzzles of atomic structure and radiation could be dealt with, and at least partially resolved, in a relatively comfortable manner. It was something of a relief that one did not have to give up the continuum, and that 'quantum' problems could be tackled using well-tried analytical methods. Schrödinger himself entertained a vision of atomic structure and 'transitions' that reduced them to manifestations of an underlying fundamental vibrational process. He deemed it both more plausible and more satisfactory to envision the atom as a system of vibrations than as a discontinuous system of 'permitted' orbits, between which electrons would unexplainably make 'jumps' that could not be

[1] "Quantisation as an Eigenvalue Problem".
[2] "Mr. Schrödinger's idea of defining the Ψ wave of a system in configuration space had first appalled me because, configuration space being purely fictitious, this conception removes all physical reality from the wave: for me, the wave of undulatory mechanics had to evolve in three-dimensional physical space" (de Broglie 1956).

pictured or analytically accounted for[1]. As to the nature of the solutions of his 'wave equation', Schrödinger could only speculate. Had it been possible, there is little doubt that he would have, like de Broglie, felt much more comfortable carrying out his programme using the resources of four-dimensional spacetime. There seemed, however, to be no possible wave-theoretical treatment of the several-electron problem that could avoid resorting to multidimensional configuration space. He suggested, without much conviction, that $\Psi\Psi^* = |\Psi|^2$ might play the part of a weight function in configuration space, quantifying spatial fluctuations of the electric charge density.

The advocates of matrix mechanics made a somewhat bitter response to the favourable reception of Schrödinger's theory. Conspicuous drawbacks of that approach were pointed out: its general formulation in multidimensional q-space, which reduced claims of 'visualisability' (*Anschaulichkeit*) to mere pretence; Schrödinger's half-baked interpretations of the Ψ solution; and the allegedly undulatory character of a framework in which a basic quantity like the velocity of the 'q-wave' appeared as a function of the mutual potential energy...of particles. The wave packets that Schrödinger had introduced to bridge the gap with the notion of localised particles[2] were also shown to have an irresistible, terribly 'uncorpuscular' tendency to spread out[3]. This made it unlikely that the orbital motion of localisable electrons within the hydrogen atom could be accounted for in terms of such packets. Schrödinger, though unyielding in his rejection of discrete 'quantum jumps', was more willing than were Pauli or Heisenberg to reconcile the two approaches. He showed that basic wave-mechanical relations actually had a counterpart in the Born-Heisenberg-Jordan formalism, provided suitable differential operators were introduced, and that a wave-mechanical equation could be consistently formulated as a matrix equation (not a major surprise once it is realised that Heisenberg's matrix elements are after all nothing but Fourier coefficients 'in disguise').

Born, at least, soon acknowledged that Schrödinger's formalism was – rather paradoxically perhaps – more suitable for a treatment of collision problems than the abstract methods he had himself contributed to develop. A suitable eigenfunction being ascribed to the incident particle, the wave-theoretical

---

[1] Schrödinger's judgment on matrix mechanics: a "rather difficult method of transcendental algebra, defying any visualisation", has often been quoted.

[2] In his *Quantisierung...* paper, Schrödinger writes: "One can try to build a wave group with relatively small dimensions in all directions. Such a wave group will then evidently follow the same laws of motion as a single representative point [*Bildpunkt*] of the mechanical system. As long as one can view it as being approximately punctual, that is, as long as one can neglect its extension with respect to the dimensions of the system's trajectory, one can view the wave group as *replacing,* as it were, the representative point".

[3] The linear harmonic oscillator, which Schrödinger had made use of to introduce his wave packets, is actually a rare and noteworthy exception.

problem came down to that of the scattering of approximately plane waves. Far from the scattering centre, the expression of the outgoing (q-space) wave took the form of a weighted sum or integral in which a 'wave' function of the angles appeared, specifying a direction of scattering. Given his predominantly corpuscular viewpoint, Born was compelled to interpret the square of the modulus of every such function as a measure of the probability of scattering in a given direction. Schrödinger's Ψ could then be regarded as a 'wavelike' propagator of such a probability, the propagation of the 'probability wave' being regulated by Schrödinger's differential equation. Born also formulated two basic theorems that were to be of central importance for the calculation of probabilities in a quantum-mechanical problem. The theorem of *spectral decomposition* states that a suitable set of eigenfunctions can be chosen for expanding linearly a given Ψ, where each eigenfunction corresponds to a possible (experimentally ascertainable) outcome. According to the second theorem, the relative phases of the coefficients in any such expansion contribute in a crucial way to the evaluation of probabilities. Born's interpretation of 'quantum' probabilities, however, was totally inconsistent:

> The dominant tendency in Born's writings is to treat the quantum probabilities, given by the square of the modulus of the wavefunction, as probabilities for *measurement outcomes* and to waffle on their premeasurement meanings, sometimes implying that they express our ignorance (unavoidable and ultimately ineliminable, owing to the uncertainty relations) regarding the objectively well-defined states of individual systems, but at other times inclining to the view that the states of individual systems are, prior to measurement, objectively indeterminate...To make matters worse, Born apparently also believed that his own 'statistical interpretation' of the quantum theory gave as complete a description of the quantum world as could be given[1].

The theorists-ideologues of the Copenhagen orthodoxy would soon radicalise the role of probability in quantum mechanics, endowing it with a quasi-ontological significance. Quantum probabilism was, as it were, the icing on the Bohrian cake: for those who had convinced themselves of the impossibility of providing any detailed account of transitions within atoms, it was but a short step to interpret Born's probabilities as both *a priori* and irreducible, and to deny the possibility of describing the actual motion of microphysical objects.
Much of the theoretical effort would, from the end of 1926, be devoted to investigating the mathematical structure of the new scheme. Dirac studied transformations of Ψ functions from the q-space to the p-space of 'conjugate' momenta, and showed that the coefficients involved in a change of basis determine probabilities of transitions through their squared moduli. A major step forward was the development of the so-called *transformation theory*, whose main purpose was to find quantum-mechanical analogues of the

[1] Deltete and Guy 1990, p.675.

canonical transformations of classical mechanics. Jordan achieved another technical *tour de force*, unifying all the previous formulations of the new theory into a tight mathematical package. Dirac's and Jordan's versions of transformation theory had all appearances of fulfilling an instrumentalist programme, for they merely supplied consistent rules for computing the probability that measuring quantity X will yield x_k, given the result y_p of measuring another quantity Y, when the linear operators associated with X and Y do not in general commute. Those like Einstein, de Broglie and Schrödinger, whose urge for a causal understanding was left unsatisfied, were more and more concerned about a drift toward formalism.

No lesser a mathematician than David Hilbert was called upon to assist in probing the mathematical structure of the new theory. However, a decisive and quasi-final step was made by von Neumann, who was the first to realise (1927) that the space of all summable and square-integrable functions is actually isomorphic to that of vector sequences involved in the discrete quantum-mechanical eigenvalue problem. Nowadays, the abstract structure underlying all the formulations of quantum theory is referred to as *Hilbert space*. Von Neumann developed a theory of linear operators and made a rigorous treatment of the eigenvalue problem, which he generalised to infinite-dimensional Hilbert space.

Heisenberg's conviction that basic traits of the quantum-mechanical formalism, e.g. the non-commutation of the operators associated with physical quantities, could not be reconciled with a causal explanation of transitions, or indeed of any process on (sub)atomic scales, was reinforced by the failure of Schrödinger's short-lived attempts at making good physical sense of the 'Ψ waves'. There was hardly any prospect of ever providing a more or less 'intuitive' picture of atomic structure and microphysical phenomena. On the other hand, if the actual purpose of physics was the proper coordination of our perceptions[1], references to space/time related conceptualisations could not be dispensed with. The applicability of classical concepts should then somehow be restricted or qualified. Unlike Bohr, Heisenberg looked for a quantitative criterion that would signal limitations to the applicability of those concepts. Whilst working on transformation theory, Dirac had pointed out that the reciprocity, associated with properties of the Fourier transform, of 'position' (q) and 'momentum' (p) in the quantum-mechanical setting implies that one's knowledge of a given numerical value of q implies maximal uncertainty regarding p (all values of p are then equally possible.) What was needed was a quantitative relationship regulating the theoretically permissible distributions of q and p values, or those of other 'conjugate' quantities.

[1] Heisenberg's early writings betray the influence of Mach, as do Einstein's analyses in his 1905 'relativity' paper.

In a 1927 paper, Heisenberg worked out a limiting case of such a relation in the form of an equality $\Delta p \Delta q = \hbar$ which, he argued, shed light on the rather enigmatic canonical commutation relation (1.8). It remained to be seen, however, whether that kind of relation was more than just a reflection of inherent mathematical traits of the theory – more precisely, of properties associated with the inversibility of the Fourier integral. Heisenberg addressed the issue by discussing the conditions of the observation of the position of an electron with a high-resolution (γ ray) microscope. This well-known *Gedankenexperiment*, which will not be reproduced here, prompted his conclusion that our attempts to acquire information about the localisation *and* momentum of a microphysical object imply an inevitable and irreducible trade-off, captured by the inequality $\Delta p \Delta q \geq \hbar$ ('uncertainty [*Ungenauichkeit*] relation'). Heisenberg's argument[1] is the prototype of a widely popularised[2] motivation of the quantum-mechanical formalism based on the idea that observing a microphysical object must inevitably disturb its course[3].

After 1924, Bohr had mostly been a keen observer of the emergence of quantum mechanics in its various guises. The last developments strengthened his long-held views. In particular, Born's suggestion to think of Ψ as a sort of probability density function must have appeared to him a natural outgrowth of his own earlier speculations about the randomness of atomic transitions. Spurred by Heisenberg's derivation of the 'uncertainty relations', the significance of which he had discussed with Heisenberg himself (and partly disagreed upon), Bohr waxed philosophical in an attempt to draw the lesson of all the efforts, achievements and debates of the past years. Unlike Heisenberg, who emphasised corpuscular aspects as those to be retained, if only in a symbolic manner, from classical physics, Bohr saw in Planck's ($E = h\nu$) and de Broglie's ($p = \frac{h}{\lambda}$) basic relations a compelling indication that the new scheme had to somehow make room for *both* classically corpuscular and undulatory aspects. These, though mutually exclusive, were none the less necessary to a proper account of the phenomenology of atomic physics. No logical contradiction was to be feared as long as the set-ups required to ascertain either of those aspects could not be simultaneously realised. The finiteness of the action quantum (Planck's constant) somehow seemed to imply that the behaviour of a microphysical system could not be generally disentangled from the effect of our attempts to observe it. The extent to which

---

[1] Somewhat surprisingly, in the same paper, Heisenberg calls upon Schrödinger's wave packets, which he had rejected when Schrödinger propounded them, to support another sweeping suggestion: that experimental interventions are required to 'substantiate' or, as it were, 'devirtualise' orbital paths.

[2] Notably by Feynman in his Caltech lectures (Feynman *et al.* 1965, sections 1-8, 2-2 and 2-6).

[3] See Brown and Redhead 1981 for a devastating critique of the disturbance-based account.

our gathering information about one aspect (e.g. 'position') would affect another (the component of 'momentum' in the same direction) was quantified by Heisenberg's inequalities. The outcome of those ruminations was delivered in a lecture at the centenary Alessandro Volta conference by Como lake in 1927. The rhetorical complexity of Bohr's presentation actually failed to captivate an audience whose technical expertise was far greater than its readiness to digest 'philosophical' lessons. However, the essentials of Bohr's ideas were, so to speak, already in the air, and it would not be long until most physicists paid lip service to the so-called Copenhagen interpretation. At the Solvay Congress that took place in Brussels one month after the Como conference, Louis de Broglie's plea for a causal alternative to 'quantum' computational methods met with utter indifference. Einstein's dissent expressed itself in a number of objections levelled at what he deemed an inherent incompleteness of quantum mechanics. Bohr responded to all those challenges, pointing out time and again that Heisenberg's 'uncertainties', when properly taken into account, would void Einstein's criticism. However, neither Einstein nor Schrödinger would completely give up. Their persistent criticism of the emerging orthodoxy culminated in the same year (1935), respectively in a paper Einstein wrote in association with two Princeton colleagues, B. Podolsky and N. Rosen, and Schrödinger's with his no less famous 'cat paradox'.

2

On 'states' and physical quantities in quantum mechanics

2.1 Quantum mechanics without 'states'?

Various notions of *state* occur in science and applied mathematics (e.g. 'finite-state automata' in computing theory). However, the word as it is used in relation to quantum theory originates in analytical mechanics. There, *state* designates a set of generalised coordinates and their derivatives, or a set of phase space coordinates (generalised coordinates and conjugate momenta). Diachronically, the change with time of the coordinates traces out a path in the relevant space, in accordance with Lagrange's or Hamilton's equations. Synchronically, the state gives a snapshot of the motion. However, it is customary to regard it also as a 'property state' of a moving object. Saying that the value of a dynamical variable lies in a certain range amounts to picking off a *subset* of the relevant phase space, namely that of classical states for which the assertion is true. Now, assuming that Hilbert space does qualify as an appropriate 'state space', a similar assertion in quantum mechanics would in contrast pick off a *linear subspace*. Can a Hilbert space ray (i.e. a one-dimensional subspace) be justifiably regarded as supplying a 'list' of properties of a moving (presumably microscopic) body?

If a given Hilbert space ray determines a quantum state, then to that ray should correspond a list of definite properties, associated with definite (eigen)values of a number of variables. However, because the algebra of idempotent variables is no longer boolean[1], properties that do not belong to the list cannot be regarded as not being the case. Taking propositions which assert properties associated with a 'quantum property state' as true will not imply that all those properties that are not associated with the property state are false; nor can we consider as false propositions that correspond to subspaces orthogonal to the ray, for then all those (infinitely many) propositions associated with subspaces that are neither orthogonal to the ray nor include it would be left with no definite truth value. Thus, the idea that properties possessed by a system are reflected in the formalism raises basic difficulties[2]. If these arise, it is because it is uncritically assumed that Hilbert space is a valid analogue of classical 'state' (phase) space, and in particular

[1] See e.g. Bub 1999, chapter 1.
[2] See Sections 4.3-4.4.

that Hilbert space vectors or Ψ functions are legitimate representatives of some property states of individual systems.

In quantum mechanics, linear combinations of Hilbert space vectors (superpositions) arguably reflect dynamical relationships that cannot be resolved using the resources of classical mechanics. However, admitting as much does not imply that such relationships warrant the interpretation of Hilbert space vectors as representatives of 'non-classical states'. Some of the pioneers of quantum theory were wary of the temptation to read too much into the new formalism. In a letter to Einstein dated 13 July 1935, Schrödinger writes: "The best response so far [to the EPR argument[1]] is from Pauli, who at least admits that the use of the word 'state' [*Zustand*] for the psi-function is quite disreputable[2]." Besides Pauli and Schrödinger, other leading figures in the development of quantum theory also expressed reservations with respect to a view of Ψ functions as representatives of 'states'.

In his Como lecture[3], Niels Bohr points out that "the definition of the state of a physical system, as ordinarily understood, claims the elimination of all external disturbances." Observation, however, implies "interactions with suitable agencies of measurement, not belonging to the system". Since, in classical physics, such interactions have very little, if any effect on our descriptive or predictive abilities, the idea of an individual system having a state is a useful and totally harmless fiction. By contrast, in the microphysical domain quantum mechanics regulates, "an unambiguous definition of the state of a system is naturally no longer possible, and there can be no question of causality in the ordinary sense of the word...[t]he space-time coordination and the claim of causality, the union of which characterizes the classical theories", should be thought of as "complementary but exclusive features of the description, symbolizing the idealization of observation and definition, respectively[4]." Such complementarity[5] precludes attributing to a given system both a localisation in space (associated classically with a definite set of generalised coordinates) and definite causal behaviour (associated in classical mechanics with a definite set of time derivatives of the generalised coordinates or with conjugate momenta). Accordingly, nothing in the quantum-mechanical formalism qualifies as the representative of the state of a physical system. Wave functions or Hilbert space vectors are, in Bohr's view, just symbolic and computational devices whose usefulness is utterly devoid of ontological significance.

---

[1] Einstein *et al.* 1935.
[2] Quoted by Fine 1986, p.74.
[3] Bohr 1928.
[4] Bohr 1928; quoted from Jammer 1989, p.366.
[5] For a lucid discussion of the evolution of Bohr's views on complementarity, see Held 1994.

In his study of collisions (1926), Max Born never regards wave functions as valid representatives of objective atomic states. Any such function provides a means to calculate the probability of a given process or 'transition', but it does not appear to connectable to a kinematics of particles. In his later writings, Werner Heisenberg[1] would also speak of Ψ functions as *probability functions*, only conceding that those functions relate, besides our incomplete knowledge, to 'potentialities' that would be inherent in such experimental situations as encountered in atomic physics.

Although he had welcomed Schrödinger's achievement, Albert Einstein soon found himself on the front line of the opponents[2] to the new 'mechanics'. He quickly realised that the new Ψ function could not be said to describe an objective physical state of an individual system unless some major concessions were made, which he deemed unacceptable. The inadequacy of Ψ functions for the purpose of describing physical states was the target of his ultimate and most famous attack on quantum theory[3]. Considering two particles **A** and **B** that previously interacted and thereby developed correlations associated with the conservation of the total linear or angular momentum, a complete description of the **A+B** pair should, Einstein contended, (i) ascribe to **A** and **B** separately a representative of their objective state. Furthermore, (ii) **B** should retain his unaltered objective state whatever occurred to the arbitrarily distant particle **A**. However, not only did quantum mechanics violate (i) in ascribing the **A+B** a single, non-factorable wave function, but contrary to (ii) different final 'states' would be ascribed to **B** depending on the outcome of measuring on **A** a relevant quantity. As Einstein wrote to Born[4], "If one abandons the assumption that what exists in different parts of space has its own, independent, real existence, then I cannot see what it is that physics is meant to describe." If realistically interpreted Ψ functions violate such a basic requirement it was enough, in Einstein's view, for denying their adequacy as representatives of states of individual bodies. If such functions relate in any way to individual systems, they do so in a fundamentally incomplete, statistical manner. Nowhere, perhaps, are Einstein's qualms about the claim that Ψ functions are representatives of states summarised in a more lucid way than in a 1952 letter to Michele Besso:

> What relation is there between the 'state' ('quantum state') described by a function Ψ and a real deterministic situation (that we call the 'real state')? Does the quantum state characterize completely (1) or incompletely (2) a real state? One cannot respond unambiguously to this question, because each measurement represents a real uncontrollable intervention in the system (Heisenberg). The real state is not therefore something that is immediately

[1] Heisenberg 1963.
[2] Deltete and Guy 1990.
[3] Einstein *et al.* 1935.
[4] Letter dated 3 March 1947, quoted by Deltete and Guy 1990, p.680.

> accessible to experience, and its appreciation always rests hypothetical (comparable to the notion of force in classical mechanics, if one doesn't fix *a priori* the law of motion). Therefore suppositions (1) and (2) are, in principle, both possible. A decision in favor of one of them can be taken only after an examination and confrontation of the admissibility of their consequences.
> I reject (1) because it obliges us to admit that there is a rigid connection between parts of the system separated from each other in space in an arbitrary way (instantaneous action at a distance, which doesn't diminish when the distance increases). Here is the demonstration:
> A system S_{12}, with a function Ψ_{12}, which is known, is composed of two systems S_1 and S_2, which are very far from each other at the instant t. If one makes a 'complete' measurement on S_1, which can be done in different ways (according to whether one measures, for example, the momenta or the coordinates), depending on the result of the measurement and the function Ψ_{12}, one can determine by current quantum-theoretical methods the function Ψ_2 of the second system. *This function can assume different forms*, according to the *procedure* of measurement applied to S_1.
> But this is in contradiction with (1) *if one excludes action at a distance.* Therefore the measurement on S_1 has no effect on the real state S_2, and therefore assuming (1) no effect on the quantum state of S_2 described by Ψ_2.
> I am thus forced to pass to the supposition (2) according to which the real state of a system is only described incompletely by the function Ψ_{12}. If one considers the method of the present quantum theory as being in principle definitive, that amounts to renouncing a complete description of real states. One could justify this renunciation if one assumes that there is no law for real states – i.e., that their description would be useless. Otherwise said, that would mean: laws don't apply to things, but only to what observation teaches us about them (the laws that relate to the temporal succession of this partial knowledge are however entirely deterministic).
> Now, I can't accept that. I think that the statistical character of the present theory is simply conditioned by the choice of an incomplete description[1].

According to Einstein, quantum mechanics in its currently accepted form would not fulfil the primary aim of a physical theory, which is to provide as complete an account as possible of the behaviour (motion) of individual objects. Fundamental physical laws cannot "consist in relations between *probabilities* for the real things, but [must consist in] relations concerning the things themselves[2]." To the extent that quantum theory has to rely on probability, it is essentially inadequate as a basic theoretical framework. Their probabilistic use suggested to Einstein that 'wave functions' referred to *ensembles* of microphysical systems. This would explain why they would be essentially inadequate for answering questions such as the decay time of a single radioactive atom. The probabilities calculated using the rules of quantum theory would merely quantify our lack of knowledge about the properties of individual members of one such ensemble[3]. Einstein occasionally mentions the ideal character of 'quantum ensembles'. This suggests that he might have thought of them as something akin to the conceptual (Gibbsian)

[1] Quoted by Bernstein 1991.
[2] Einstein, quoted by Fine 1986, pp.100-101.
[3] Einstein's insistence on relating quantum-theoretical probabilities to ensembles suggests that his 1916 introduction of a priori probabilities was in his view nothing but a makeshift, that should be discarded when a satisfactory dynamical framework would be available.

ensembles of classical statistical mechanics. As to whether he thought of anything more specific, there is room for speculation. Would Planck's early (1912) derivation of the correct expression for the energy of the quantised harmonic oscillator count as a step toward such an ensemble view? By averaging over a phase-randomised ensemble of harmonic oscillators at thermodynamic equilibrium, Planck had succeeded in deriving the $\frac{\hbar\omega}{2}$ value of zero-point energy per harmonic oscillator. This was achieved using classical statistical reasoning (with a crucial reference to the non-negativity of probability), some thirteen years before Heisenberg's quantum-mechanical treatment of the same problem. Another similar example is a simple derivation of the expression $j(j+1)\hbar^2$ of the squared modulus of the total angular momentum on the basis of assuming random orientation in a statistical ensemble and computing classical averages. Feynman[1] ascribes this "interesting and somewhat surprising" result to the fact that equiprobability as such means fundamentally the same thing whether in a classical or a 'quantal' context. This is why applying a simple averaging procedure e.g. to quantised momenta can *sometimes* lead to the correct expression. All the same, it is seldom actually possible "to guess the correct formula by using the classical calculation[2]". Unlike Post[3], who advocates an ensemble interpretation of quantum theory on the very slim basis of the above two instances, it is unlikely that Einstein could have taken such coincidences so seriously as to even start thinking of them as evidence for 'quantum ensembles'. Presumably, the reason why he did not care to be more explicit about his view of such ensembles is that he was convinced that quantum mechanics was a mere expedient[4], which would eventually have to be discarded in favour of a genuinely fundamental physical theory:

> If the statistical quantum theory does not pretend to describe the individual system (and its development in time) completely, it appears unavoidable to look elsewhere for a complete description of the individual system; in doing so it would be clear from the very beginning that the elements of such a description are not contained within the conceptual scheme of the statistical quantum theory[5].

Only an instrumentalist could think of the current framework as the last word, but this was an option Einstein rejected vigorously, as it was incompatible with his lofty conception of scientific endeavour[6]. Instead, a radical refoundation

---

[1] Feynman *et al.* 1965, section 34-7.
[2] *Ibid.*
[3] Post 1995.
[4] Deltete and Guy 1990, p.679.
[5] Letter to Michele Besso, quoted in Schilpp 1970.
[6] "If that were so [i.e. if quantum mechanics or any other allegedly fundamental physical theory were to be accepted on purely instrumentalistic grounds] then physics could only claim the

of theoretical physics would be called for, in which the appearance of the quantum would find as much of a justification as the constant **c** does in the special theory of relativity[1]. In this alternative and hopefully final framework there would not – *could* not – be any room for probability or statistics at a fundamental level. In a letter to Louis de Broglie dated 15 February 1954, Einstein recalls how such a goal led him to his final solitary endeavour, away from the mainstream:

> En vérité, je suis, exactement comme vous, convaincu qu'il faut chercher une substructure, une nécessité que la théorie quantique actuelle cache habilement par l'application de la forme statistique. Mais depuis longtemps je suis convaincu qu'on ne pourra pas trouver cette substruture par une voie constructive en partant du comportement (empirique) des choses physiques connues, car le saut conceptuel nécessaire dépasserait les forces humaines. Ce n'est pas seulement par la futilité de nombreuses années d'efforts que je suis arrivé à cette opinion, mais aussi par mon expérience en théorie de la gravitation. Les équations de la gravitation pouvaient être découvertes seulement sur la base d'un principe purement formel (la covariance générale) c'est-à-dire sur la base de la conviction que les lois de la nature ont la plus grande simplicité logique imaginable. Comme il était évident que la théorie de la gravitation ne constitue qu'un premier pas vers la découverte de lois générales de champ les plus simples possibles, il me semblait que d'abord cette voie logique devait être poursuivie jusqu'à la fin avant de pouvoir espérer arriver également à une solution du problème quantique[2].

Dirac may well have been the first of the most prominent theorists of quantum mechanics to make extensive use in his writings of the word 'state', and intend it in its problematic acceptation. The idea of particles being 'in states' featured prominently in his 1925 study of the quantum gases[3]. However, it finds its full expression in his influential 1930 treatise[4]. There, setting up a new kind of mechanical framework is seen as the necessary outcome of the recognition of "a limit to the gentleness with which observations can be made[5]". We are thereby led to redefining the size of an object in terms of its susceptibility to

---

interest of shopkeepers and engineers; the whole thing would be a wretched bungle." (letter to Schrödinger, quoted from Przibram 1967, p.39).

[1] See Section 3.1.

[2] Einstein 1989, p.249: "Actually, I am, just like you, convinced we have to look for a substructure, a necessity that the present quantum theory skilfully hides by its application of statistical form. However, I have been convinced for a long time that we won't be able to find that substructure in a constructive way, starting from the (empirical) behaviour of known physical things, for the necessary conceptual leap would be beyond human powers. I have reached that conclusion, not only because of the futility of many years of effort, but also through my experience in gravitation theory. The equations of gravitation could be discovered only on the basis of a purely formal principle (general covariance),that is on the basis of the conviction that the laws of nature have the greatest imaginable logical simplicity. Since it was evident that the theory of gravitation is but a first step towards discovering general laws of field that are the simplest possible, it seemed to me that this logical path first had to be followed to the end before one could hope to also get to a solution of the quantum problem."

[3] Darrigol 1986.

[4] Dirac 1958.

[5] *Ibid.*, p.4.

disturbance: "If the object under observation is such that the unavoidable limiting disturbance is negligible, then the object is big in the absolute sense and we may apply classical mechanics to it. If, on the other hand, the limiting disturbance is not negligible, then the object is small in the absolute sense and we require a new theory for dealing with it[1]." The possibility of maintaining causal connections would be restricted to undisturbed systems and find its expression in suitable differential equations. A key difference with classical mechanics is that the new framework should also make the connection between some 'initial' state of affairs (before observation) and whatever can be the case after a disturbing observation. On account of the "consequent indeterminacy in the results of our observations[2]", the theory would allow us in general only to calculate the probability that any subsequent observation yields a particular result.

Dirac introduces the *principle of superposition* as "one of the most fundamental and drastic[3]" of a new set of "laws of nature" whose necessity arises from the inevitability of disturbance[4]. His intention is to provide insights into the way the two notions of probability and disturbance connect. Two basic examples open the discussion, both of which involve monochromatic light. In the case of a beam of light passing through a crystal of tourmaline, Malus's law gives the fraction of outgoing light as a simple (co)sine-squared function of the angle between the direction of polarisation of the incident light and the optic axis of the analyser. The description of light as "made up of photons[5]" makes it sound like Dirac favours a corpuscular picture. As long as the incident light is polarised parallel or perpendicular to the optic axis, a photon-based interpretation is straightforward: each photon can be thought of as either stopped (absorbed) or going through. The interpretation requires thinking of each individual photon as possessing a '(linear) polarisation' attribute, or equivalently as being "in a certain *state of polarization*[6]". The relative proportions of ingoing photons polarised parallel or perpendicular to the optic axis would then be reflected in the probabilities that one photon passes or fails to pass through the crystal. Repeated detections of outgoing photons should allow us to check that the relative proportion of outgoing photons agrees with the fraction calculated using Malus's law for a given angle away from the optic axis.

---

[1] *Ibid.*
[2] *Ibid.*
[3] *Ibid.*
[4] To the extent those laws relate to the effect on *our* predictive capacities of the disturbances caused by *our* ways and means of observation, the reference to nature itself raises questions.
[5] *Ibid.*,p.5.
[6] *Ibid.*,p.5.

The fate of an "obliquely polarised" photon is not *a priori* so easily settled. Dirac dismisses from the outset as unscientific[1] any attempt at describing what takes place inside the tourmaline crystal between the time an individual photon gets in and the time 'it' comes out: "only questions about the results of experiments have a real significance and it is only such questions that theoretical physics has to consider[2]". This creates a tension with Dirac's insistence on "preserv[ing] the individuality of the photon in all cases[3]", which is resolved by contending that it is *our observation* of the outgoing beam that "force[s] the photon entirely into the state of parallel or entirely into the state of perpendicular polarization[4]." On account of the uncontrollable character of the accompanying disturbance, which final state an individual photon would "jump[5]" into could not in principle be predicted in any better way than probabilistically.

Let us concede that it is sensible to think[6] of each photon in a beam as being "located somewhere in the region of space through with the beam is passing[7]". The photon can also be ascribed in the direction of the beam a momentum whose magnitude is proportional to the frequency (Einstein's photo-electric law). Saying that the photon "is in a definite *translational state*[8]" is just another way of expressing our having "information about the location and momentum of a photon." As in the previous example, Dirac acts as a realist in respect of photons, although it is ambiguous whether saying an individual photon is in a translational state is to be regarded as asserting an objective (transient) property of the photon or merely lays some pseudo-ontological veneer on our assessment of an experimental situation (as suggested by Dirac's use of the words "know" and "information"). Besides, ascribing a photon a momentum that is proportional to frequency amounts to conceding undulatory features that cannot obviously be reconciled with a granular view of light. In fact, saying that "the photon is partly in one beam and partly in the other[9]" is tantamount to denying that the photon could be a particle in any acceptable sense of the word. Dirac does not appear to be concerned with such difficulties. "The essential point", as he sees it, "is the association of each of the translational states of a photon with one of the wave functions of ordinary wave optics." Consistency problems that follow from postulating some kind of wave-particle 'duality' need not worry us, for "[t]he nature of this association cannot be

---

[1] "outside the domain of science" (Dirac 1958, p.6).
[2] *Ibid.*
[3] *Ibid.*, p.6.
[4] Dirac 1958, p.7.
[5] *Ibid.*
[6] Dirac writes, more affirmatively, that we "know" it (Dirac1958, p.7).
[7] Dirac 1954, p.7.
[8] *Ibid.*
[9] *Ibid.*,p.8.

pictured on a basis of classical mechanics, but is something entirely new.[1]" In the translational case, beam splitting is achieved using an interferometer. There again, the upshot is that one should somehow think of the state of each photon as split up into two components, its final state being a "superposition of the two translational states associated with the two components.[2]" Attempting to determine the energy of one of the components would precipitate a sudden change of state, "due to the disturbance in the translational state of the photon which the observation necessarily makes[3]." As a result of such a disturbance, it is "impossible to predict in which of the two beams the photon will be found. Only the probability of either result can be calculated from the previous distribution of the photon over the two beams.[4]" The interferometer example is also used to introduce the idea that a superposition of two translational states somehow involves interference. As in the polarisation example, no detailed explanation of such interference can be given at an 'interphenomenal' level. Dirac argues that preserving conservation of energy implies that "[e]ach photon then interferes only with itself[5]" – a very odd and uninformative statement if there is one. Inscrutable though such 'self-interference' may be, on should be reminded that "the main object of physical science is not the provision of pictures, but is the formulation of laws governing phenomena and the application of these laws to the discovery of new phenomena. If a picture exists, so much the better; but whether a picture exists or not is a matter of only secondary importance[6]." What really matters is that the "principle of superposition of states" participates in ensuring the self-consistency and predictive efficiency of the new theory.

Considering any (sub)atomic-size systems, elementary or composite, engaged in any kind of interaction, "[t]here will be various possible motions of the particles or bodies consistent with the laws of force. *Each such motion is called a state of the system*[7]." In classical mechanics, instantaneous numerical values of the coordinates of the dynamical system in the appropriate configuration space, and the corresponding velocity components, are both necessary and sufficient to determine an entire motion. Quantum mechanics, if it is based on the recognition that the disturbances suffered by 'small' systems cannot be arbitrarily reduced or ignored, "puts a limitation on the number of data that can be assigned to a state[8]." As a result, a quantum-mechanical

---

[1] *Ibid.*.
[2] *Ibid.*
[3] *Ibid.*
[4] *Ibid.*
[5] *Ibid.*,p.9.
[6] *Ibid.*,p.10.
[7] *Ibid.*,p.11 (italics added).
[8] *Ibid.*

> "state of a system may be defined as an undisturbed motion that is restricted by as many conditions or data as are theoretically possible without mutual interference or contradiction. In practice the conditions could be imposed by a suitable preparation of the system, consisting perhaps in passing it through various kinds of sorting apparatus, such as slits and polarimeters, the system being left undisturbed after the preparation. The word 'state' may be used to mean either the state at one particular time (after the preparation), or the state throughout the whole of time after the preparation. To distinguish these two meanings, the latter will be called a 'state of motion' when there is liable to be ambiguity[1]."

Dirac's decision to skip the explicit reference to motion in the synchronic use of the word 'state' makes it all too easy for that word to be read, incorrectly, as denoting a formal characterisation of a given system's 'being' at a particular instant. It takes no more than such a slip to fall head-on into the infamous paradoxes and conundrums that continue to plague the conceptual debate about the quantum theory.

The postulated expression of a state as a superposition of various other states is "a mathematical procedure[2]" that "is always permissible, independent of any reference to physical conditions, like the procedure of resolving [the mathematical expression of] a wave into Fourier components.[3]" The procedure is justified in so far as it is useful. Acknowledging its usefulness should take precedence over any attempt to interpret relationships between 'superposed' states, which "cannot be explained in terms of familiar physical concepts[4]." We obviously "cannot in the classical sense picture a system being partly in each of two states and see the equivalence of this to the system being completely in some other state. There is an entirely new idea involved[5]". As to what the 'idea' might consist in, Dirac treads dangerously close to attempting an ontological reading of superpositions: "When a state is formed by the superposition of two other states, it will *have properties* that *are* in some vague way [?] intermediate between those of the two original states and that *approach more or less closely to those of either of them* according to the greater or less 'weight' attached to this state in the superposition process[6]." Faced with a conceptual difficulty, Dirac typically reverts to the safety of an operationalistic position. Dispelling any temptation to give an ontological reading of the puzzling kind of intermediacy that is supposedly exhibited in superpositions, he settles the matter by asserting that "[t]he intermediate character of the state formed by superposition thus *expresses* itself through the *probability* of a particular result for an observation *being intermediate between the corresponding probabilities* for the original states,

---

[1] *Ibid.*,pp.11-12.
[2] *Ibid.*,p.12.
[3] *Ibid.*
[4] *Ibid.*
[5] *Ibid.*
[6] *Ibid.*, p.13 (italics added).

not through the *result* itself being intermediate between the corresponding results for the original states[1]." The contentious matter of whether definite attributes – and if so, which – can be said to be possessed by a system 'in' a superposed state is conflated, without further ado, with the trivial observation that measurement results are never 'seen in superposition' (whatever that could mean). Questions that one might be tempted to raise regarding the definiteness of 'possessed values' between observations are brushed aside as metaphysical issues of no concern to the physicist. The principle of superposition, which reflects the recognition of an irreducible susceptibility of microphysical objects to disturbance, "demand[s] indeterminacy in the results of observation[2]". Period. As a result, developing a probabilistic framework with built-in linear (?) features would be legitimate...and inescapable.

Whether or not Dirac is to be regarded as *the* originator of the current notion of 'quantum state' is of secondary importance. The fact remains that a view of quantum mechanics as being fundamentally about the specification, evolution in time and manifold transformations of 'quantum states' emerged along with the first abstract formulations of the principles of quantum mechanics. After Dirac, von Neumann's authoritative monograph[3] (1932) also greatly contributed in promoting a state-based conception of quantum theory. It is also not coincidental that von Neumann should have been led in his treatise to the first expression and discussion of the 'measurement problem', for this and related objectification[4] problems are inevitable consequences of conceiving of Hilbert space vectors as representives of physically meaningful states of individual objects.

Physics, like other fields, has its own specific terminology in which, besides newly coined words, new use is found for older words borrowed from earlier stages in its development. 'Momentum' or 'Hamiltonian' in quantum mechanics are instances of the latter. The use of the word *state* in quantum theory is another, albeit one whose potential for breeding confusion should not be underestimated. Renewing the use of older terminology in an entirely different setting may be convenient, perhaps even desirable. However, terminological decisions can have potentially harmful effects, because expectations associated with the former uses of words may interfere in a destructive way with our efforts to make sense of a new framework. This is all the more so in quantum mechanics, the theoretical underpinnings of which are still obscure and debated. The decision, if it is insufficiently motivated (e.g. beyond superficial analogies), is likely to bias one's views, with lots of unpalatable consequences. For example, *entanglement* may not, as it was intended by

---

[1] *Ibid.* The whole sentence appears in italics in the text.
[2] *Ibid.*,p.14.
[3] Von Neumann 1955.
[4] See Section 4.4.

Schrödinger, just bring to mind the purely mathematical fact of non-factorability implied by ⊗-product composition, but conjure up, in its association with a 'state' view, the far more problematic vision of individual systems being tangled up in a mysterious way.
Here as elsewhere, the later Wittgenstein's warnings[1] should be heeded: a terminological decision is not the result of an empirical discovery, forced upon scientists by their practice. It is only worth what conventions are, and may be misguided if whatever motivates it suffers from a basic lack of insight into the new domain to which it is applied. It is hardly needed to say that amending such a decision in the 'quantum state' case is far from sufficient, for it clearly falls short of providing answers to basic conceptual issues raised by the emergence and mathematical structure of quantum theory. Nevertheless, it is advisable to heed the attitude of some of its founding fathers (regardless of their personal appraisal of the theory) and abstain from using a word that is too heavily laden with ontological connotations to be so readily – uncritically – accepted.

2.2 The representation of physical quantities

Elementary (Galilean-relativistic) quantum theory cannot, expect perhaps in a loosely historical sense, be regarded as *deriving* from classical mechanics. On the other hand, it is most important to be aware of any presuppositions, or of any regulative structures, that may be shared by those apparently very dissimilar frameworks. Traditional invocations of quantum-classical 'correspondence' are powerless to give any insight into the most significant features of quantum theory. Thus, regarding commutators as quantum analogues of the Poisson brackets provides little information about their central role in quantum mechanics. Moreover, quantum-theoretical quantities do not all have a classical counterpart – *spin* is a case in point. By contrast, focussing on symmetry-invariance principles goes a long way towards elucidating any common structure. Foremost among such principles are those that relate to basic requirements of descriptions in space and time, for the satisfaction of such requirements takes precedence over any detailed account of whatever interactions systems are engaged in.
Starting from the actual endpoint of the historical development of the theory, let us assume we have a theoretical framework involving Hilbert space vectors and linear operators, as they are used for the computation of probabilities in quantum theory. The question then arises of how to specify, out of all linear operators that are available a priori, the set of self-adjoint operators that are

[1] Wittgenstein 1960, p.23.

to represent physically significant quantities. Although his initial approach to deriving the 'quantum conditions' hinges on the shaky idea of classical-quantal correspondence, Dirac anticipates later group-theoretical motivations of the formalism in his 1930 treatise[1]. How kets and linear operators are affected by simple displacements can, he argues, provide "new insight[2]" into the quantum conditions. Relations involving 'states' (represented by kets) or dynamical variables (represented using operators) are expected not to be affected if those states and variables are all displaced in the same way: "any symbolic equation between kets, bras, and dynamical variables must remain invariant under the displacement of every symbol occurring in it, on account of such an equation having some physical significance which will not get changed by the displacement.[3]" Considering a displacement through a distance δx in the x-direction and letting δx tend to zero, Dirac introduces an infinitesimal displacement operator d_x (in the x direction) that satisfies[4] the relation $xd_x - d_x x = 1$. Now, this is basically the same commutation relation as that between the component p_x of momentum and the x cartesian coordinate. Therefore, the component p_x of total momentum may be *defined*, with the introduction of Planck's constant, as a MLT^{-1}-dimensional expression of the displacement operator along x : $p_x = i\hbar d_x$. That different components of momentum have vanishing commutators then directly reflects the basic geometrical fact that infinitesimal displacements in different directions are commuting operations. Similarly, requiring invariance of linear relationships under time displacement, i.e. under a resetting of the zeroes of clocks, leads to an equation for the evolution in time of arbitrary kets: $i\hbar\frac{d}{dt}|u\rangle = \mathbf{H}(t)|u\rangle$, which coincides with Schrödinger's time-dependent equation (the linear operator **H** then plays the role of a 'time displacement operator'). Dirac justifies his conceiving of **H** as a representative of the total energy of the system by asserting that "the [special] theory of relativity puts energy in the same relation to time as momentum to distance.[5]" Referring to Einsteinian relativity in the process of deriving a basically Galilean-relativistic formalism is, to put it mildly, incongruous. Dirac, like many others, appears not to have been aware at the time that the relations in question reflect the invariance-conservation relationships that form the subject matter of Noether's theorem (1918). That theorem connects conserved quantities like total momentum and energy to the requirement of invariance under space and time displacements (translations) respectively, in the 'pre-Einsteinian' framework of Lagrangian

---

[1] Dirac 1958.
[2] *Ibid.*, p.99.
[3] Dirac 1958, p.101.
[4] *Ibid.*, pp.102-103.
[5] *Ibid.*, p.110.

mechanics. The resilience of Noetherian relationships in the face of the most drastic forms of theory change (from Galilean to Einstein-relativistic, from classical to 'quantal') testifies to their fundamental character. Their survival in the quantum setting certainly is the essential reason why "[t]he laws of conservation of energy, momentum, and angular momentum hold for an isolated system in the Heisenberg picture [of quantum mechanics], as they hold in classical mechanics[1]."

According to Wigner's theorem, a symmetry transformation of Hilbert space vectors amounts to the action on those vectors of a(n) (anti)unitary operator. Requiring invariance under a symmetry transformation implies the existence of a unitary projective (Hilbert space) representation of the relevant group. If the symmetry group is continuous (Lie group), then for each of its one-dimensional subgroups there must exist a self-adjoint operator that plays the role of an infinitesimal generator for that subgroup (Stone-Naimark theorem). The operator representative of the group can be written as a product $\prod_k \exp(i\alpha_k A_k)$, where $\{A_k\}$ is a set of self-adjoint operators and each of the real parameters α_k parameterises a one-dimensional subgroup. As Jean-Marc Lévy-Leblond points out,

> "it has become customary for physicists to describe with the same notation the elements of the Lie algebra and their operator representatives, as well as the Lie bracket of two elements and its representation by the commutator of two infinitesimal generators. It is now clear that the set thus obtained of self-adjoint operators furnishes good candidates to the title of physical properties of the quantum system. Indeed, all these operators have simple meanings as they are associated with invariance under some one-parameter group (e.g. time translations); they can be given their usual names by comparison with, and extension from, their classical analogues (e.g. it is well known in classical mechanics that time translations are generated by the energy, or Hamiltonian, of the system). In this approach, the reason for the particular importance of the commutator structure of the operators representing the physical properties becomes particularly transparent, since it only expresses their Lie algebra characteristics, that is, ultimately, their transformation rules. Further physical properties of interest may then be built in the enveloping algebra, that is as functions of the infinitesimal generators; this can often be done by relying once more on symmetry arguments, for instance the transformation rule of the properties sought after (which, once more, will be expressed through commutation relations). Among those, a crucial role is given to the invariant properties, that is the elements of the centre of the enveloping algebra, since they characterize the representation of the group; they are diagonalized for irreducible representations, and their (eigen)values thus label the elementary physical systems[2]."

---

[1] *Ibid.*, p.115. The sentence appears in italics in the original.
[2] Lévy-Leblond 1974.

A key to the efficiency of both classical and quantum-mechanical schemes lies in the fact that invariance under such transformations is almost invariably, if only tacitly, assumed as part of our ordinary experience and our experimental as well as theoretical practice of physics. Sure enough, we can have no a priori guarantee that any such invariance will remain valid as physicists venture into hitherto unexplored territories. Nevertheless, it is a rather weak requirement that is independent of the particulars of the interactions systems may be engaged in. Moreover, recognising that classical and quantum mechanics, despite obvious dissimilarities, rely on the same basic transformational properties, which all come down to setting up a suitable space-time Galilean background for dynamics, provides a common ground for gauging the effectiveness of 'correspondences' that would otherwise sound implausibly accidental.

A basic transformation group operating in classical mechanics is the Galilei group. This group is, in one dimension, generated by three one-dimensional subgroups: that of space translations; that of time translations; and that of pure Galilean transformations (associated with relative motion). The group is non-abelian: "jumping into a train running at 60 miles per hour and waiting for 5 minutes takes you 5 miles away from where you would be if you had first waited for 5 minutes and then jumped into the train[1]." Constructing the relevant quantum-mechanical Lie algebra is done using standard infinitesimal methods. The Lie bracket or commutator formed with the Galilean and time generators does not generally vanish (it is equal to the space translation generator), thereby exhibiting the non-abelian character of the Galilei group. A little algebra leads to the possible projective representations of the group, and in particular the expression of the continuous exponents. Within a suitable extension of the group by an abelian one-dimensional 'phase group' (an extension based on no assumption of physical significance), a realisation of the irreducible representation of the Lie algebra follows, without resorting to any analogy with classical mechanics[2]. The ensuing expressions of the generators, diagonal in P (the space translation generator), supply the basic quantities required for a quantum-theoretical treatment of the dynamics of an elementary system.

Switching from (momentum) P-space to 'configuration space' requires introducing a 'position operator' X that correctly transforms under a space translation. Applied to an infinitesimal displacement, this group-theoretical requirement implies that X satisfies the canonical commutation rule $[X,P] = iI$, where I is the unit operator and the imaginary factor *i* is introduced to guarantee that the generators are Hermitian. Another requirement is that the X operator transforms in a correct manner under pure Galilean transformations.

---

[1] *Ibid.*

[2] see Lévy-Leblond 1974 for details.

This in turn implies that X is actually proportional to the Galilean transformation generator. The characteristic mass *m* of the system then emerges as the inverse of the proportionality constant. An appropriate function Ψ of space and time coordinates can now be introduced via Fourier transform. Owing to its definition and to the form of the position and momentum operators, the Ψ function automatically satisfies the Schrödinger equation. Its role is to translate into space-time coordinates the fundamental relation $H = \frac{P^2}{2m} + U$, derived on purely group-theoretical grounds, between the Hamiltonian (energy) operator H (which corresponds to the time displacement generator) and the momentum P (which corresponds to the space displacement generator). An alternative route consists in starting from the Schrödinger equation in configuration space, working out the way it must transform to be Galilei-invariant, deducing the projective form of the Galilei group representation and from there the operatorial expression of the relevant dynamical quantities (Hamiltonian, angular momenta etc.). This route has the disadvantage of relying explicitly on an equation whose initial motivation and physical meaning remain obscure besides the probabilistic interpretation of its solution.

Extending the formalism beyond the free particle case requires taking interactions into account, leading to a modified form for **H**. Owing to the symmetry-breaking effect of an external field, no thoroughly group-theoretical approach is now applicable. However, requiring that instantaneous Galilean transformations exist and preserve the validity of the Galilean-relativistic velocity addition law can drastically restrict the form of the modified Hamiltonian. This is tantamount to requiring that no discrimination be possible instantaneously between the free and the interacting cases, and can be formalised as a constraint on H since the velocity operator depends on the Hamiltonian. The generator of pure Galilean transformations remains the same as in the free case, but the Hamiltonian operator becomes a function of X and P: $H = \frac{1}{2m}[P - A(X,t)]^2 + U(X,t)$, which is the familiar form of the Hamiltonian for a system under the influence of a scalar (U) and a vector (A) potential.

A three-dimensional generalisation has to include transformations under rotations. The three-dimensional Galilei group is a ten-parameter Lie group, whose Lie algebra is exhausted by giving the generator H of time translations, the generators P_x, P_y, P_z of spatial translations along three orthogonal axes, the three generators of pure Galilean transformations and the generators J_x, J_y, J_z of rotations about the three axes. As in the one-dimensional case, looking for the unitary irreducible representations associated with elementary

systems leads to an invariant $U = H - \frac{P^2}{2m}$ that can be readily interpreted as the internal energy of the system. A position operator X being introduced as a direct generalisation of the one-dimensional case, an operator L can then be constructed from X and P, in analogy with the vector product $\mathbf{r}\times\mathbf{p}$, which defines the classical angular momentum. Owing to the Lie brackets involving the generators J_x, J_y, J_z, the expression of the cartesian components of L warrant regarding it as an angular momentum operator[1]. The difference $S = J - L$ between L ('external' or 'orbital' angular momentum) and the vector operator J (total angular momentum) whose components are J_x, J_y and J_z is a linear operator whose components S_x, S_y, S_z commute with all the generators of the Lie algebra of the Galilei group, but not among themselves (they satisfy $[S_x, S_y] = iS_z$ and circular permutations thereof). On the other hand, S^2 is an invariant of the Lie algebra (it commutes with all the generators and with the components of S). Likewise, the definition and properties of S warrant its interpretation as an internal angular momentum. The dimensionless value of S^2 in any irreducible unitary representation is shown to be of the form s(s+1), where the *spin* number s can be an integer or a half-integer, and can be used as a 'quantum number' for labelling a system type. Spin is thus seen to emerge from constraints imposed on the linear representative framework of quantum theory by the structure of the three-dimensional Galilei group (involving translations, rotations and inertial motions). Changes of reference frame i.e. space-time orientation imply a modification of dynamical quantities like energy or angular momentum, but do not affect S.

Summing up,

- Neither the mathematical structure of quantum theory nor tentative analogies with classical mechanics warrant the interpretation of Hilbert space vectors or Ψ functions as representatives of *states* of individual systems. This claim does not amount to asserting that these vectors and functions are completely devoid of physical significance, but rather that their theoretical significance has to be thoroughly reconsidered.

[1] Goldstein (1980, p. 419) remarks that "one of the earliest references to the classical Poisson brackets for angular momentum appears to be the 1930 treatise by Born and Jordan on *Elementäre Quantenmechanik*....while the general change of a vector function under rotation has long been familiar in quantum mechanics, until very recently about the only reference to its classical version was in the famous thesis of H.B.G. Casimir, *Rotation of a Rigid Body in Quantum Mechanics*, 1931, p.30."

- *Given* the Hilbert space structure which underlies the rules for the calculation of probabilities, average values and other quantities of operational interest, the *form* of the relevant physical (observable) quantities, with their mutual relationships (commutation relations in particular), can be derived as the result of imposing group-theoretical constraints associated with elementary space, time and inertial transformations.

3

Quantum mechanics as a principle theory

3.1 The Special Theory of Relativity: a prime instance of a principle theory

In Newton's mechanics, it is – as it turns out, incorrectly – assumed that the relative time of occurrence of any two events and the order in which they are (can be) seen to occur are fixed in an absolute way. Those persistent misconceptions of time and simultaneity obviously find their source in our ordinary experience: do not we seem to share a single present? Isn't the 'rate at which time flows' the same for all of us regardless of our relative motion? Einsteinian relativity disposes of all these prejudices, replacing them with operational definitions of such notions as 'time interval', 'simultaneity', 'length of a moving rod' etc. Instead of assuming that we have an a priori 'clear and distinct' understanding of space and time, the Special Theory of Relativity (STR) adjusts our conceptions of time and space to carefully performed readings of our measuring instruments (typically clocks and rulers). As a result, observers who bear different space-time relationships to two events **A** and **B** need not agree as to whether the two events occur at the same time, **A** precedes **B** or vice versa.

It was subsequently realised[1] that Einstein's postulate of the invariance of the speed of light in all reference frames can actually be dispensed with. The basic change-of-frame (Lorentz) transformation of STR can be derived[2] on the basis of the sole 'principle of relativity', without having to postulate some invariant velocity, or to refer in any way to electromagnetic waves. According to the relativity postulate, there exists an infinite continuous class of reference frames – so-called *inertial* frames – in which the mathematical expressions of the physical laws have the same form (claiming that such inertial frames exist is tantamount to denying the existence of absolute motion, hence of absolute rest). The purpose of a theory of relativity is to provide rules of transformation for relating the expression of any given law in one frame to its expression in another of the same class.

Deriving the Lorentz transformation requires making a small number of basic assumptions about space, time and the composition of inertial transformations:

---

[1] E.g. Ignatowsky 1911.

[2] Lee and Kalotas 1975, Lévy-Leblond 1976.

- *Homogeneity* of space and time: the transformation properties of a space-time interval do not depend on where the endpoints of the interval are located in the chosen frame.

This assumption implies that inertial transformations are linear, and that motions obtained from rest by an inertial transformation are uniform motions (their velocity is constant in the 'target' frame).

- *Isotropy* of space: there are no directions in space that have a privileged status (all possible orientations are physically equivalent).

Considering, to simplify, frames such that one spatial coordinate x only is non-trivially affected by an inertial transformation, spatial isotropy implies that if the sets (x,t) and (x',t') are related by the transformation we seek, then so are (-x,t) and (-x',t'). From this requirement, the seemingly obvious result that if S' moves relative to S with velocity v, and S moves relative to S' with velocity u, then u = -v can be derived.

- The inertial transformations have a *group* structure[1]: identity corresponds to null relative velocity; an inverse inertial transformation from S' to S exists for every inertial transformation from S to S' (the parameters of the two transformations are shown to be the opposite velocities v and -v); and the composition of two successive inertial transformations $S_1 \rightarrow S_2$ and $S_2 \rightarrow S_3$ is an inertial transformation $S_1 \rightarrow S_3$.

As a result, if S_2 moves relative to S_1 with velocity v_{12}, and S_3 moves with velocity v_{23} relative to S_2, there must exist a single velocity v_{13} that relates the two frames S_1 and S_3. Moreover, the frames being inertial, the sought coordinate transformation must be of the same form as that between S_1 and S_2, and that between S_2 and S_3. Given the other (homogeneity and isotropy) assumptions, the composition law for velocities has to be of the form $v_{13} = \frac{v_{12} + v_{23}}{1 + \alpha v_{12} v_{23}}$. The change-of-frame transformation also involves a linear coefficient whose expression is a function $(1 - \alpha v^2)^{-1/2}$ of the relative velocity v of the two frames involved. Three cases now arise, depending on the sign of the real parameter α, or whether that parameter is equal to zero. Although a choice of negative α is permitted a priori, this choice can be ruled out by rejecting the possibility that two velocities in the same direction add up to a

[1] The linearity of the transformation reflects the fact that the translation group, associated with arbitrary shifts in space and time origins, is an invariant subgroup of the full transformation group.

velocity in the opposite direction[1]. Equivalently, $\alpha < 0$ can be rejected on the grounds that the sign of a space-time interval between events should not change under inertial transformations[2] (this is a sensible requirement since a transformation is nothing but a change of perspective). We are then left with two choices: $\alpha = 0$, which yields the familiar Galilean transformation, or $\alpha > 0$, which corresponds to the Lorentz transformation provided we let $\alpha = \frac{1}{\kappa^2}$. The constant κ then plays the role of a universal limiting velocity. Naturally, nothing in the assumptions leading to the 'Lorentz' i.e. $\alpha > 0$ transformation implies that any such upper limit to 'physical[3]' velocities is realised in actual fact in the physical world as we experience it. Assuming no universal limiting speed amounts to letting κ tend to infinity, thus reverting to the $\alpha = 0$ Galilean option as the sole candidate for a theory of relative motion (v_{13} is then equal to $v_{12} + v_{23}$, in accordance with our ordinary experience). However, it so happens that a finite value of κ appears to be realised in the propagation of electromagnetic waves, of which visible light is an instance. It is indeed a most remarkable coincidence (?) that light, which propagates with the highest *known* speed, *also* travels at the maximum permissible speed[4].

In hindsight, it seems to be sheer luck that physicists could hit upon a telltale sign (when recognised!) that the Galilean option was in fact the wrong one. In that context, it is little surprise that the relativistic issue was initially mistaken for a difficulty with electromagnetism. Indeed, no less a figure than Henri Poincaré, who went so far as to obtain the correct form for the 'Lorentz' transformation and realise its group structure, fell prey to that delusion[5]. Einstein's position as an outsider was certainly beneficial to his coming to realise that the problem had fundamentally been misconceived. However, he did not himself completely disentangle the key issues from references to electromagnetism, as we can see from his basing his theory of relativity on *two* distinct postulates rather than a single relativistic one. The transition from the Galilean to the Lorentz-Einstein transformation would have been both conceptually more radical and satisfactory, had Einstein fully realised that the nature and properties of light only contingently contributed to the issue of

---

[1] Lee and Kalotas 1975.

[2] Lévy-Leblond 1976.

[3] It should be clear that the value supplied by κ is an upper bound to the relative velocities of two inertial reference frames, *not* the velocity of 'anything' observed *in* any given reference frame (values of the latter are not constrained in any way by STR).

[4] As far as we presently know. If electromagnetic waves were ever found to propagate with a (very) slightly lower velocity than *c*, this fact would in no way invalidate the principles of STR, since those are seen to be independent of any assumption regarding the nature and properties of light.

[5] Poincaré, who is sometimes incorrectly credited as 'co-discoverer' (or inventor) of Special Relativity, was also held back by his own philosophical inclinations; see Miller 1998, pp. 174-177, p. 240.

relativity. In the first part of the 1905 paper, the problem of synchronising a set of standard clocks using light rays gives the latter an importance electromagnetic radiation need not have in that context: any kind of signal propagating at a fixed and uniform speed in the rest frame of its source would qualify. However, the propagation of light met all of the prerequisites for Einstein's discussion of what should count as proper synchronisation of clocks, simultaneity etc.

Mathematically, STR boils down to the statement that expressions of the laws of physics are invariant under the Poincaré group (inhomogenous Lorentz group). Transformations of that group, which preserve Minkowskian distance, can actually be regarded as compositions of reflections in straight lines, in much the same way any transformation of the plane that preserves Euclidean distance can be defined as a composition of at most three reflections[1]. One fundamental difference between the two cases is that Minkowski space(time) exhibits preferred directions (so-called 'null lines' which, incidentally, appear to coincide with the paths of light rays), whereas the Euclidean plane has none. The Galilean and Lorentzian options each have consequences that must be subjected to empirical test. Thus, in contrast to the Galilean case, the Lorentzian implies the existence of a non-vanishing 'residual' energy associated with any physical body in its rest frame. Experimental evidence gathered throughout the twentieth century is known to lend overwhelming support to the $\alpha > 0$ option, i.e. to a framework (STR) in which $\kappa = c$.

Summarising,

> **S1.** STR is grounded on a single relativity principle, which embodies a small set of compelling assumptions regarding basic properties of the space-time framework used for coordinating events.
>
> **S2.** Amongst the two possible types of ensuing inertial transformations, the 'Lorentz-Einstein' one (LE) features a new fundamental constant κ, usually noted c, which yields an upper bound on the attainable velocity of a moving body. The other case is the ordinary Galilean transformation, which formally connects to LE as its $c \to \infty$ limit.
>
> **S3.** On the LE option the classical notion of a velocity as a quantity that can be augmented ad infinitum by addition is not valid. Nevertheless, the ordinary notion may be retained for

[1] Dunn 1981.

everyday purposes or for treating problems involving much lower speeds than the value of *c* in the same units.

S4. The new fundamental constant happens to correspond, to a high degree of accuracy, to the empirical value of the speed of propagation of electromagnetic waves, hence that of light, in vacuo.

S5. LE is borne out by a very wide range of experimental tests (none of which, significantly, overlaps with our ordinary perceptual experience).

STR is perhaps the best instance of what Einstein referred[1] to as a *principle theory*: a theory whose foundations consist of a small set of general basic statements, observational consequences of which can be tested experimentally. Einstein argues that thermodynamics is such a theory, since it "seeks by analytical means to deduce necessary conditions, which separate events have to satisfy, from the universally experienced fact that perpetual motion is impossible[2]." That no putatively perpetual motion has ever been found to withstand critical examination does not, strictly speaking, warrant the claim that the impossibility of any form of *perpetuum mobile* is an experienced fact. In the STR case, caution regarding the empirical status of the basic assumptions is all the more required that the most obviously empirical of the two postulates in the original (1905) formulation – that of the invariance of the velocity of light – is in fact unnecessary, and alien as it stands to the very motivations and theoretical significance of STR. Principle theories like STR are certainly empirical in some of their motivations, but they also give expression to conditions of comprehensibility that one inclined to Kantism might wish to call 'transcendental' – an aspect that Einstein and those with a realist bent are more likely to overlook or just ignore.

To principle theories, Einstein opposes *constructive* theories. The kinetic theory of gases provides a classical instance of such a theory, which is characterised by a typically 'bottom-up', entity-based approach to the explanation of phenomena. Elementary building blocks (e.g. atoms or molecules) being postulated, higher-level processes, like diffusion, are shown to emerge from the behaviour of the simple underlying entities (e.g. molecular motion). Those who think of such reductive efforts are virtually synonymous to the aim of science might well dismiss principle theories as mere constructive theories *manquées*. However, the conceptual economy and clarity of STR, once it is properly grounded on a single principle, speaks for itself: it is as compelling as a universal framework can be. So is the General

[1] Einstein 1954.
[2] Einstein 1954, p.228.

Theory of Relativity (1916), despite a tendency to reify the space-time metric that leads to some thorny problems[1]. Constructive approaches arguably fare best wherever imagery and metaphors can be heuristically helpful. In contrast, a principle approach appears to be suited to those regulative frameworks whose principles have to be complied with by theories with quite different subject matters (e.g. gases, solids, stellar objects etc). Thus, quantum mechanics sounds like the very kind of theoretical framework one might expect to be formulated as a principle theory (Einstein himself might well have wished it to be so). Instead, its current status is that of a conceptually baffling though highly successful compendium of computational techniques, which has nothing of the conceptual lucidity of STR.

3.2 Some steps toward a principle-based account of quantum mechanics

If quantum mechanics is to be regarded as a principle theory, a basic set of compelling postulates has to be found, from which some basic equation must directly follow. The equation, which may or not be Schrödinger's, should embody the essentials of quantum mechanics without any 'non-standard' addition or omitting any aspect of the current framework. The basic postulates from which the theory can be derived should be statements of conditions that, like isotropy for STR are expected, or justifiably required, to hold. These conditions need not refer to space and time but might e.g. express requirements to be fulfilled by a set of generalised coordinates in configuration space. On the other hand, in accordance with Einstein's views, any considerations of a probabilistic or statistical nature appear to be quite out of place at that fundamental level.

If, following the STR example, some new universal constant ξ appears in the process of deriving the theory, then ξ is expected to be simply related (e.g. proportional) to Planck's constant. The heuristic value of the idea of classical-quantal correspondence also suggests that classical mechanics should, in analogy with (S3), be formally recovered in the appropriate limit (say, as $\xi \rightarrow 0$). This would amount, conceptually, to the relaxation of a structural limitation implied by the set of grounding postulates. Strictly speaking, classical and quantum mechanics would correspond to the options $\xi = 0$ and $\xi \neq 0$ respectively. The two frameworks would therefore stand in much the same kind of relationship as the Galilean and the Lorentz-Einstein options in the context of working out appropriate inertial transformations. This would have totally escaped the attention of theorists because the appearance

[1] E.g. the 'hole argument'; see Butterfield 1989, Earman 1995.

of Planck's constant did not give any clue as to what kind of structural constraints could lead to the derivation of a quantum-mechanical framework. In contrast, the relatively prompt and neat 1905 resolution of the pre-Einsteinian crisis was facilitated by two circumstances: that the puzzle obviously had something to do with relative motion and with what seemed (misleadingly) to be a puzzling feature of the propagation of light.

Now if, in close analogy to (S2), the hypothetical constant ξ is to express some fundamental limitation set to physical processes – or to our description thereof – then this limitation must be clearly identified. On account of the difficulties encountered since 1900 in trying to make sense of Planck's quantum, the essential significance of ξ is unlikely to be so easily read off the relations implied by the grounding postulates. If ξ must be interpreted as an 'action quantum', we can hardly expect the said limitation to be so straightforward as the recognition of an upper limit to 'physical' velocities. Be it only dimensionally[1], *action* turns out to be ubiquitous in mechanics, classical or quantum, Galilean or Einstein-relativistic. Most significantly, products of the members of symmetry parameter/conserved quantity pairs associated through Noether's theorem: time×energy, position (coordinate)×momentum component, (angular value×)angular momentum all have the dimensions of an action. The selfsame 'conjugates' appear in Heisenberg's inequalities and they also form the scalar (time, energy) and vector (position, momentum) contributions to 4-vectors. Such is the central role of action in the variational formulation of mechanics[2] that it has been claimed that "*action*, not *energy*, is the basic concept in general dynamics.[3]" In view of the greater abstractness of this notion – far greater, indeed, than that of velocity, be it in STR – any fundamental limitation involving it is bound to be subtle and not so easily amenable to our intuition. It is likely to point to deeply structural constraints, which would account in part for their having been overlooked in the actual historical development of quantum theory (another obstacle being the intellectual climate that prevailed in the days that saw the emergence of the theory).

To give an idea of what a fundamental limitation relating to 'action' might be, consider a bundle of configuration space curves, each of which corresponds to a physical or realisable motion, i.e. to a motion that gives rise to stationary action between two configuration endpoints **A** and **B**. These motions differ from one another by the choice of the constants appearing in the expansions of the solutions q and p of Hamilton's canonical equations. If any one motion is to be unambiguously determined, no point can belong to more than one

---

[1] The dimensions of the magnitude 'action' are mass×(length)^2×(time)$^{-1}$.

[2] Synge and Griffiths 1959 (chapter 16) ; Goldstein 1980.

[3] Synge and Griffiths 1959, p.435 (italics in original).

curve. One further condition of 'coherence[1]' is imposed on that set of physical motions: the definite action integral $I_A^B = \int_{t_A}^{t_B} L dt$, where $L = \sum_k (p_k \dot{q}_k - H)$ should not depend on the path followed in connecting **A** and **B** (the condition is equivalent to $\oint L dt = 0$ along any closed curve in the relevant configuration+time space). The condition I_A^B = *constant* defines a surface that corresponds to the geometrical locus of the **B** points (a different choice of **A** merely adds a constant to the integral). Setting I(q,t) = *constant* for a given value of t determines a surface that 'propagates' as a sort of wave front as time varies. Propagation does not take place in ordinary space. Rather, at some initial time, all physical motions that are compatible with the initial conditions and physical constraints are, as it were, generated at once. From then on, all configuration paths unfold with velocities that span the range from 0 to ∞, giving rise to an ideal expanding 'wave'. Evidently, only *one* among all these jointly generated motions is actually instantiated and ever actually observed (in real space).

To make it less abstract, let us focus on the simplest instance of action wave propagation, namely the free motion of a point mass *m* in one dimension. The generalised coordinate is chosen as the cartesian coordinate q = x. The Lagrangian then reduces to the constant kinetic energy: $L = \frac{1}{2} m \dot{x}^2 = E$, where $\dot{x}$ denotes the velocity. If the physical motions we are interested in are those that start from x = 0 at t = 0, the expression of the action is then $I(x,t) = \int_0^t L dt = \frac{1}{2} m \frac{r^2}{t}$, where r = |x| is the distance to the origin, and $\dot{x} = \frac{r}{t}$.

The equation $\frac{1}{2} m \frac{r^2}{t} = \sigma$ = *constant* defines a surface of action that propagates in configuration (x-)space with a velocity that is exactly half that of the point mass. There is thus no question – in case one had any doubt – of picturing the particle as 'surfing' on the wave. Action waves have no phenomenological significance beyond their providing a means of describing *in toto* the time evolution of *all* physical motions that are *a priori* compatible with the initial set.

Let I(q,t) = σ and I(q',t') = σ' be the equations of two constant action surfaces Σ and Σ'. If **M**∈Σ and **M'**∈Σ', then since the integral $I_M^{M'} = \int_{t_M}^{t_{M'}} L dt$ between any two points **M** and **M'** within a coherent set of motions does not depend on the path connecting the two points, then $I_M^{M'} = \Delta\sigma = \sigma' - \sigma$ is the same for all the

[1] Synge and Griffiths 1959, p.455.

motions in the set. In other words, the increment in action between the two surfaces is not a characteristic of the actual (observed) motion of the system; it is a characteristic of the coherent set ($\Delta\sigma$ can take a priori any value between 0 and ∞). In our example, whereas Σ reaches a distance r from the origin after a time $t = \frac{mr^2}{2\sigma}$, Σ' does so at the earlier time $t' = \frac{mr^2}{2(\sigma+\xi)}$. Assuming that ξ is very small, the time difference is $\Delta t = t - t' = \frac{mr^2}{2\sigma^2} = \frac{2\xi}{m\dot{x}^2}$, so that $E = \frac{\xi}{\Delta t}$. Letting Δt be the period of action waves, a crest of which passes through **M'** at t' and through **M** at t, the frequency is $\nu = \frac{1}{\Delta t}$ and $E = \xi\nu$. This suggests identifying ξ with Planck's constant *h*, leading to $E = h\nu$. Tentatively generalising the conclusion to more complex dynamical systems, let us postulate that the constant *h* is the smallest increment in action between two action wave crests. Admitting such a bound on $\Delta\sigma$ exists, any physical motion should satisfy the condition $|\Delta\sigma| \geq h$. A sketchy connection to quantum mechanics can be made as follows: the form $\Psi(x,t) \propto e^{\frac{i}{\hbar}(Et-p_x x)}$ of a typical solution to Schrödinger's one-dimensional equation, where $Et - p_x x$ has the dimension of an action, suggests writing a general solution as $\Psi(x,t) \propto e^{\frac{i}{\hbar}S(x,t)}$. The basic dynamical equation $i\hbar\frac{d\Psi}{dt} = H\Psi$ can then be written $\left(\left(\frac{\partial S}{\partial t}\right) + H\right)e^{\frac{i}{\hbar}S} = 0$ in terms of the action function S(x,t). The equation is generally satisfied if $\frac{\partial S}{\partial t} + H = 0$, i.e. if S satisfies an analogue of the classical Hamilton-Jacobi equation, where the 'phase' S (in units of $\hbar$) of the complex exponential plays the role of Hamilton's principal function.

An important aspect of the relation between S and Ψ is that it is many-to-one. The Ψ function is invariant under the transformation $S \rightarrow S \pm nh$, where *n* is a natural number: the Ψ representation does not discriminate between S and S' if S' – S is a multiple of Planck's constant, where S' – S is just the increment $\Delta\sigma$ in classical action and *h* the lower bound on such an increment we tentatively introduced. Since no non-zero lower bound is implied by classical mechanics, the question arises of what factors might determine its occurrence. Alternatively, one may decide to take the existence of such a bound as a basic postulate on which quantum physics should be grounded. However, the existence of a non-zero lower bound on the action wave increment is

everything but self-explanatory. Indeed, it is enough to compare it with the usual interpretation of the universal constant of STR to see how much more compelling and 'natural' the latter is. Another major drawback of this approach is that, if the introduction of an action quantum can be motivated as we just did, it is far from obvious how we should go about from there to deriving Schrödinger's equation. For that reason, we shall leave action waves aside and concentrate instead on the above-mentioned analogy of form to the Hamilton-Jacobi equation.

Writing the solution of Schrödinger's equation in polar form $\Psi(r,t) = R(r,t)e^{\frac{iS(r,t)}{\hbar}}$ and separating into real and imaginary parts yields the two equations:

$$\frac{\partial S}{\partial t} + \frac{(\nabla S)^2}{2m} + V = \frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} \qquad (3.2.1)$$

and

$$m\frac{\partial R}{\partial t} + \nabla R.\nabla S + \frac{1}{2}R\nabla^2 S = 0 \qquad (3.2.2)$$

Equation (3.2.2) can also be written $\frac{\partial R^2}{\partial t} + \nabla.(\frac{R^2\nabla S}{m}) = 0$, i.e. as $\frac{\partial \rho}{\partial t} + \nabla.J = 0$ with $\rho(r) = |\Psi(r)|^2$ and $J = \frac{R^2\nabla S}{m}$. Since this is just the kind of continuity equation one encounters in hydrodynamics, a 'fluid' metaphor has been tentatively applied to Ψ: the function would describe an assembly of non-interacting point masses whose dynamics, subject to the potential V, would be akin to that of a non-viscous fluid. However, since its propagation would not take place in real three-dimensional space but in configuration space, the fluid in question has to be regarded as an essentially abstract theoretical construct (no less abstract, indeed, than waves of action are). Accordingly, its density ρ is usually taken to have no other meaning than that of a probability density of occurrence while J, such that $\nabla.J = \frac{i}{\hbar}[\Psi^*(H\Psi) - (H\Psi)^*\Psi]$, is referred to as the 'probability current density'. Equation (3.2.1) is more directly relevant to addressing the genealogy and characteristics of quantum dynamics. It is remarkable in that, were its right hand side equal to zero, it would be just identical with the classical Hamilton-Jacobi equation, whose solution is Hamilton's principal function S. It is, indeed, as if all potentially quantum-mechanical effects were embodied in the non-vanishing right hand side. This observation is actually the starting point of an alternative approach to quantum theory, first introduced by Louis de Broglie[1] and, some twenty-four years later,

[1] De Broglie 1928.

rediscovered by David Bohm[1]. The parallel with classical mechanics can be made even more transparent by writing (3.2.1) in the form

$$\frac{\partial S}{\partial t}+\frac{(\nabla S)^2}{2m}+V+Q=0, \qquad (3.2.3)$$

where $Q=-\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R}$. With the subsidiary 'guidance' condition $p = \vec{\nabla} S$ and letting $E = -\frac{\partial S}{\partial t}$, this equation becomes an expression of the conservation of energy $E = \frac{p^2}{2m}+V+Q$, provided Q is regarded as an additional 'quantum' potential, nonexistent in classical mechanics, and which is held entirely responsible for the difference between quantum-mechanical and classical behaviour. Equation (3.2.1) generates a set of one-parameter curves that Bohm and his followers identify with definite trajectories of particles, subject to the guidance condition[2]. Equation (3.2.2) ensures that the computation of probabilities remains in agreement with quantum statistics: provided the trajectories of particles happen to be distributed (which raises the question of why they should) with a probabilistic weight R^2 over a set of initial conditions, the weighted distribution of those trajectories will match the predictions of quantum theory in its standard formulation.

It is an attractive feature of Bohm's approach that we don't seem to be fated to abandon trajectories. Indeed, considering the variations of the quantum potential might seem to afford a smooth transition from classical (vanishing Q) to fully Bohmian trajectories. Besides, given that (3.2.1) follows from Schrödinger's equation by a simple substitution, the issue of providing an interpretation of the quantum potential cannot be evaded. The Q potential may tentatively be regarded as a new form of energy, which would be required to guarantee conservation of energy in all those processes whose typical actions are of the same order of magnitude as Planck's constant. Clearly, the very form of Q implies a radical departure from classicality. If it might seem that a restoration of intelligibility along classical lines is just round the corner, the impression is nothing but an illusion. After all, equation (3.2.3) owes its very form to its being *derived from* the Schrödinger equation, and the Q potential is a function of the R-field i.e. the amplitude of the Ψ field. This has two (related) consequences. Firstly, the quantum potential inherits non-separability features from the Ψ function. In particular, if Ψ is a 'many-particle' function that cannot

---

[1] Bohm 1952; Holland 1993.

[2] If the system is conservative and if coordinates are Cartesian, the classical kinematical relation $p = m\dot{q}$ follows from the guidance relation and the identifcation of $\dot{q}=\frac{1}{m}\nabla S$ with the streamline velocity of the probability current.

be factored into separate individual contributions, then neither can the Q potential on a given particle, which will generally depend on the instantaneous positions of all the other particles. As a result, on an ontic reading of the formalism, it is as if one particle's interacting externally may affect instantaneously at a distance ('non-locally') the trajectories of all the other particles that make up the system. Such non-locality connects to a remarkable feature of the quantum potential: it is independent of the amplitude of the R-field (multiplying R by a constant factor clearly has no effect on Q). This implies that the motion of particles may be significantly affected by that potential even in regions where the value of R appears to be negligible.

The non-classical properties of the quantum potential (especially form dependence) prompted Bohm to deny that it could sensibly be regarded as a mechanical potential. Whatever underlies all manifestations of quantum behaviour could not, he claimed, be reduced to any mechanical picture; the concessions that would be required for maintaining such a picture would defeat its very purpose. The quantum potential should rather be thought of as the surface mathematical expression of some inherently complex and barely conceivable 'subdynamics'. Form dependence and non-locality eventually led Bohm and Hiley to suggest that the R-field might act as an 'organising field', and to introduce the idea of *active information*[1]. The word 'information' is taken here literally and etymologically as referring to a capacity to shape or 'bring order' into something, where that 'something' might be a collection of individuals or some underlying indivisible wholeness[2]. At any rate, the Bohm-Hiley active information has little, if anything, to do with the well-established and technical concept of information, where the latter, being "information for us[3]", is devoid of ontological, 'observer-' or abstraction-free significance. In particular, information in the ordinary sense is not something that can be said, except in a loose metaphorical manner, to be 'carried by' or be 'transferred between' objective physical entities.

Bohm and Hiley conceive of multiparticle entanglement in terms of particles sharing a "common pool of information[4]". An entangled Ψ function would 'encode' such information – understood as a mysterious organising power – that is supposed to determine the behaviour of what we ordinarily conceive (perceive?) as separate entities ('particles' etc.). This suggestion clearly calls for an entirely new kind of ontology, where particles would be higher-level emanations of what Bohm refers to as 'the implicate order[5]'. The common pool of information idea would provide a clue as to why Ψ functions are more

[1] Bohm and Hiley 1993; Hiley 1999.
[2] Bohm 1983.
[3] Bohm and Hiley 1993; Maroney and Hiley 1999.
[4] *Ibid.*
[5] Bohm 1983.

generally definable within a higher-dimension configuration space, reducing to three-dimensional 'physical' space only in the case of a single independent particle. The irreducibility of the configuration space representation would be ultimately due to the behaviour of N particles being generally co-determined by the pool of information they share in an indissoluble manner, and which an entangled Ψ function represents at the level of quantum theoretical description. Disentanglement would mean that the particle(s) concerned would no longer have access to that pool. As a result, their behaviour, while still regulated by a wave function, would be amenable to a space-time account.

Those theorists[1] who stick to Bohm's original (1952) approach tend to downplay its holistic aspects, but the challenge of making sense of (3.2.1) remains. From our perspective, none of the variants of Bohm's endeavour is satisfactory, for the reason that all of them rely on Schrödinger's equation being given *a priori*. By contrast, a principle-based refoundation of quantum mechanics should have that equation, or some alternative 'master' equation, *emerge* from the imposition of a small set of compelling postulates. That obtaining a square-integrable solution of the Schrödinger equation is a prerequisite for getting R, S, and for deriving the kinematics (Bohmian trajectories) clearly suggests – an impression confirmed by the essential role of equation (3.2.2) – that Bohm's approach is actually rooted in the standard approach, with its particular emphasis on the squared modulus of the wave function. This and, more generally, the *a posteriori* character of the Bohmian enterprise make it an unlikely candidate for the kind of refoundation of quantum physics that we are looking for.

The Bohmian view takes advantage of the favourable circumstance that (3.2.3) differs from the classical Hamilton-Jacobi equation only by an *additional* term. This suggests focusing on that term, interpreted as a new kind of potential – a term that would, as it were, be the culprit for all 'quantum effects'. But it is by no means obvious that anything similar to the Hamilton-Jacobi equation should actually arise from Schrödinger's by just writing the Ψ function in polar form, nor that the outcome should differ from the classical equation by an additional 'corrective' term (after all, say, the Lorentz transformation does not obtain by adding terms to the analytical expression of the Galilean transformation). Last but not least, the guidance condition is suggestive of a 'trivialising map' whereby the current set of phase space coordinates is transformed to a set that is constant in time[2]. What is oddly missing is the dual of that canonical transformation. It is as if the separation of Schrödinger's equation into (3.2.1) and (3.2.2) would point toward some sort of generalisation of, or yet-to-be alternative to Hamilton-Jacobi theory, but

[1] E.g. Goldstein 1996,1998; Holland 1993.
[2] Holland 2001a.

without giving more, as it were, than 'half a hint'. The dependence on S of the quantum potential also involves higher derivatives of S, in contrast to the classical Hamilton-Jacobi equation. What is the structural function and theoretical necessity of such higher derivatives?

Substituting the wave function, in the same polar form, into the one-dimensional stationary Schrödinger equation $-\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2}+V(x)\Psi=E\Psi$ gives rise, again, to two equations[1]:

$$S'^2-2m(E-V)=\hbar^2\frac{R''}{R} \qquad (3.2.4)$$

and

$$2R'S'+RS''=0 \qquad (3.2.5)$$

From (3.2.5), $R\propto(S')^{-\frac{1}{2}}$ which, inserted in (3.2.4), yields

$$\frac{1}{2m}(S')^2+(V-E)=\frac{\hbar^2}{4m}\left[\frac{3}{2}\left(\frac{S''}{S'}\right)^2-\left(\frac{S'''}{S'}\right)\right]. \qquad (3.2.6)$$

The 'Quantum Hamilton-Jacobi Equation' (3.2.6) essentially differs, as in Bohm's case, from its classical analogue by its non-vanishing right hand side. However, what is remarkable about the right hand side of (3.2.6) is that it is proportional to the so-called 'Schwarzian derivative' of S. This third-order non-linear differential equation, whose solution is a real function S, can be used[2] as a replacement for Schrödinger's stationary equation. The initial conditions for the wave function determine, up to a constant of integration, a unique 'characteristic function' S, which is the generator of the motion for a single trajectory as a function of time. $\frac{\partial S}{\partial x}$ then corresponds to the conjugate momentum[3]. Each trajectory is sufficient to specify the wave function (there is no need to invoke an ensemble of such trajectories). On the other hand and

---

[1] $S'=\frac{\partial S}{\partial x}$, etc.

[2] The WKB (Wentzel-Kramers-Brillouin) approximation method (see Messiah 1961, Merzbacher 1970) consists in expanding S in a power series in $\hbar^2$: $S=\sum_{k=0}^{\infty}\hbar^{2k}S^{(k)}$, keeping only the zeroth-order terms. The result $\left(\frac{\partial S}{\partial t}\right)^2\approx\left(\frac{\partial S^{(0)}}{\partial t}\right)^2=2m[E-V(x)]$ then lends itself to straightforward integration.

[3] The conjugate momentum is not identical with the ordinary linear momentum. It may be worth recalling that, in the Hamiltonian formulation of classical mechanics, even if the generalised coordinate q is Cartesian, the conjugate momentum need not be identical with $m\dot{q}$ (it is not if the dynamical problem involves a velocity-dependent potential).

for purely mathematical reasons, the representation afforded by (3.2.6) does not always resolve the wave function into single trajectories.

Equation (3.2.6) has made appearances in a number of suggested alternatives to standard quantum theory, e.g. in Nelson's stochastic quantisation programme[1]. Amongst the relatively rare and unfashionable studies of (3.2.6), it is worth mentioning some well-crafted papers by Floyd[2]. Besides its own merits, Floyd's representation is of particular interest in that, unlike Bohm's, it has been set in a wider perspective by the seminal work of Faraggi and Matone, to which the next section is mainly devoted. These authors show how one can actually dispense with Floyd's need to rely, just as Bohm does, on the prior availability of the Schrödinger equation. Instead, equation (3.2.6) is itself shown to be a direct consequence of assuming the universal validity of a single unambiguously mechanical and, in a precise sense, 'relativistic' principle.

Floyd considers (3.2.6) *in toto* rather than as an expression of a 'correction' to classicality (which, as already mentioned, is a tendency of Bohm and his followers). Remarkable aspects of his approach are his identification of microstates[3] of the Schrödinger equation (which do not correspond in general to the $L^2(R)$ solutions of the standard account) and his derivation of the kinematics directly from (3.2.6). In the bound case, he shows that any given trajectory generated as a solution of (3.2.6) is consistent with a unique eigenfunction of energy as determined by Schrödinger's equation. In that sense, (3.2.6) affords a finer resolution of the 'bound state' problem than does the usual focus on wave functions. Nowhere in the description is any randomness involved. Nevertheless, the probability rule of the standard account cannot be dispensed with given our general inability (whether real or dogmatically assumed) to resolve wave functions into distinct microstates. According to Floyd[4], the use of probabilities in quantum theory is made necessary by the incomplete specification of particular microstates that the Schrödinger equation affords. On the other hand, any Floydian trajectory that corresponds to a microstate is sufficient to specify one solution of that (Schrödinger's) equation.

In Floyd's 'trajectory representation of quantum mechanics', as he calls it, and in contrast to Bohm's, the conjugate momentum is not identified with the mechanical momentum. No guiding Ψ field is required either, since motion is completely determined by (3.2.6). Moreover, Floydian trajectories, though consistent with the requirement that probability should be conserved, are not distributed in accordance with the wave function density. Another major

---

[1] Nelson 1985.
[2] E.g. Floyd 1996, 2000.
[3] Floyd's terminology.
[4] Floyd 2000.

difference with Bohm's interpretation is that Floyd's trajectories, which are more numerous than Bohm's, may cross (this may happen given an eigenstate of the Schrödinger equation that corresponds to different microstates). In the case of unbound systems, although there are no microstates and each trajectory is uniquely determined by a set of initial conditions, Floydian and Bohmian trajectories correspond to different space-time paths – the kinematical equations are different. Floyd's trajectories also differ from Feynman paths in that the generator of the motion can be regarded as an appropriate variant of Hamilton's characteristic function, whereas Feynman assumes a classical propagator.
Because of the difficulty of obtaining analytical solutions of (3.2.6), closed form solutions are yet to be obtained for dimensions higher than 1 or beyond simplified pseudo-2-dimensional situations. Another technical drawback is the considerable difficulty of time-independent perturbation theory: because the right hand side of (3.2.6) does not vanish, the task is bound to be more challenging than classical canonical perturbation theory, which is itself much more complicated than standard quantum-mechanical perturbation theory. Since perturbative treatments based upon (3.2.6) may yield results at variance with those of standard calculations[1], this could provide and incentive to test Floyd's predictions (if anyone ever cares to do so), hence the validity of his alternative approach.

3.3 Quantum mechanics from a single principle

In a recent paper, A. Faraggi and M. Matone[2] have shown in a very detailed and conclusive manner that quantum mechanics can be derived from a *single* principle. More precisely, Schrödinger's equation is shown to be an indirect consequence of requiring that dynamical laws comply with a so-called 'equivalence' principle. That principle is relativistic in so far as it can be traced back to reconsidering the special character of the rest frame in classical mechanics. The new approach provides insight into the solutions of the Schrödinger equation and sheds more light on the duality of canonical variables. No satisfactory refoundation of quantum mechanics as a 'principle theory' will be achieved in a cheap way. The Faraggi-Matone programme outlined in this section[3] is clear testimony to the level of technical sophistication required, in stark contrast e.g., to STR.

---

[1] Floyd 2000.
[2] Faraggi and Matone 2000.
[3] The reader is referred to Faraggi and Matone 2000 for all the mathematical details and proofs. For easy reference, the notation used here only differs slightly from the one used in that paper.

Canonical transformations are those transformations from a (p,q) set of conjugate coordinates and momenta to another (P,Q) that leave the form of Hamilton's equations of motion invariant. Dynamics is preserved provided that

$$p\dot{q} - H = P\dot{Q} - K + \frac{dF}{dt},$$

where H(q,p,t) and K(Q,P,t) are the old and the new Hamiltonian, F the generating function[1]. Setting $p = \frac{\partial F}{\partial q}$ and $P = -\frac{\partial F}{\partial Q}$ leads to $H + \frac{\partial F}{\partial t} = K$.

The *Hamilton-Jacobi equation* $H + \frac{\partial S}{\partial t} = 0$ obtains by requiring that K vanish (F is in then usually noted S and referred to as Hamilton's *principal* function). As a result, the formerly independent q and p variables have become dependent – in particular, $p = \partial_q S(q,Q,t)$. Integration yields Q and P, and inversion of the coordinates the solution to the dynamical problem. If the Hamiltonian can be neatly separated into a kinetic and a potential term as $H = \frac{p^2}{2m} + V(q,t)$, the Hamilton-Jacobi equation becomes

$$\frac{1}{2m}\left(\frac{\partial S}{\partial q}\right)^2 + V + \frac{\partial S}{\partial t} = 0 \qquad (3.3.1)$$

Assuming that the potential does not depend on time and writing $S = S_0 - Et$, with E the energy, the *characteristic* function a.k.a. 'reduced action' S_0 satisfies the *classical stationary Hamilton-Jacobi equation* (CSHJE):

$$\frac{1}{2m}\left(\frac{\partial S_0}{\partial q}\right)^2 + W(q) = 0 \qquad (3.3.2)$$

where $W = V(q) - E$. The dependence on time is implicit. Time parameterisation is introduced at a later stage, usually by identifying the conjugate momentum p with the mechanical one $m\dot{q}$, or through Jacobi's theorem: $t - t_0 = \partial_E S_0$.

Now consider a similar problem to that of setting up equation (3.3.2), with the difference that the relation $p = \partial_q S_0$ between reduced action and canonical momentum is introduced from the outset, instead of treating q and p as independent variables. The question raised by Faraggi and Matone is whether a coordinate transformation $q \rightarrow q'$ exists that maps a dynamics characterised by $S_0(q)$ to another characterised by $S_0'(q')$, so that they can be regarded as equivalent, e.g. in the sense that $S_0(q) = S_0'(q')$. The old and the new set of coordinates are independent variables, and dynamics is completely characterised by the functional dependence of S_0 on q and that of S_0' on q'.

[1] F is a function of t and any of the pairs (q,Q), (q,P), (p,Q), (p,P) considered as independent variables.

Since the systems are deemed arbitrary, there should exist a coordinate transformation mapping $S_0(q)$, corresponding to a system with $W \neq 0$, into the reduced action that corresponds to the free system with vanishing energy, i.e. with $W^0 = V(q^0) - E = 0$. This is reminiscent of Hamilton-Jacobi theory, although coordinate transformations only are considered, the transformation of p being induced by that of S_0 through $p = \partial_q S_0$.

As it turns out, requiring that $S_0^A(q^A) = S_0^B(q^B)$ be well-defined irrespective of the choice of systems A and B is incompatible with classical mechanics. Conflict can be traced back to a peculiarity of the rest frame: from (3.3.2), a free particle (point mass) with vanishing energy has a constant reduced action, which no coordinate transformation can connect to a non-constant function. On the other hand, such a coordinate transformation makes sense from the point of view of an observer in motion relative to both A and B, since both systems are then ascribed non-constant S_0's. Thus, since the connection we are looking for is possible only between systems, neither of which is described by a constant reduced action, the requested equivalence under coordinate transformations requires choosing a frame in which no particle is at rest. Formally, the classical impossibility can be overcome by considering time-dependent coordinate transformations, which Hamilton-Jacobi theory itself cannot accommodate. In classical mechanics, where space and time are considered on an essentially different footing, the 'passage' of time is extrinsic to a system and indifferent to whatever value spatial coordinates may be set to. Therefore, it is always possible to reduce to the free system at rest by setting $q' = q - f(t)$. An example of such a (clearly invertible) transformation is the transformation $q \rightarrow q' = q - \frac{1}{2}gt^2$, which reduces the motion of a point mass in a gravitational field (expressed by Newton's law as $m\ddot{q} = mg$) to that of a free particle ($m\ddot{q}' = 0$), where the latter naturally includes the particle at rest. Thus, time parameterisation provides a means of compensating for the 'privilege' of the rest frame in classical Galiean mechanics.

Faraggi and Matone's programme consists in setting up a dynamical framework in which no such 'classical' restrictions are imposed, tacitly or not, on the existence of locally invertible coordinate transformations. The 'new' dynamical framework is thus required to be such that an arbitrary system can be reduced to the null W^0 case. This suggests the existence of some energetic contribution, independent of the form of the potential V, that could prevent reduced action to ever be constant, thereby depriving the rest frame of any suspicious privilege. Naturally, such a contribution is not to be enforced in an *ad hoc* manner. Rather, a modification of the CSHJE (3.3.2) should result from consistently satisfying the demands of a basic 'structural' principle.

Whilst classical mechanics never treats canonical variables on a truly equal footing, the involutive nature of the Legendre transformation, which is implicit in the very definition of the generating function, gives rise to explicit p-q duality (Hamilton's equations of motion are p-q symmetrical, up to sign interchange). This duality can be made manifest by introducing the generating function T_0, which stands in the same relation to q as S_0 does to p, i.e. $q = \partial_p T_0$. Duality means that a system with given W can be described in two equivalent manners, associated with S_0 and T_0 respectively. Writing Hamilton's principal function $S = S_0 - Et$, and $T = T_0 + Et$, where $T = q\partial_q S - S$, then

$$\frac{\partial S}{\partial t} = -\frac{\partial T}{\partial t}. \qquad (3.3.3)$$

Equation (3.3.3) ensures that the S-T Legendre duality remains stable under time evolution.

Amongst all possible interchanges of the S_0 and T_0 pictures according to $q \to \alpha p$, $p \to \beta q$, that with the highest symmetry is such that $\alpha = \beta$, which leaves unchanged the functional dependence of q and p. Stability under time evolution of the Legendre relationship between S and T then requires that $\alpha\beta = -1$, consistently with the minus sign in (3.3.3). Consequently, $\alpha = \beta = \pm i$: the occurrence of imaginary numbers follows from requiring maximum symmetry in the context of Legendre duality, and that the duality be temporally stable. This is in fact as far as the imaginary factor occurring in the relation between S_0 and solutions of the Schrödinger equation (to be derived later) can be traced back.

The duality of S_0 andT_0 can be given a 'geometrical' expression by noticing that S_0 obtains as a result of a Möbius transformation of T_0. Möbius coordinate transformations[1] $q \to q' = \frac{Aq+B}{Cq+D}$ are such that $q'\sqrt{p'}$ and $\sqrt{p'}$ are linear combinations of the solutions to two second-order linear differential equations[2]

$$\left(\frac{\partial^2}{\partial s^2} + \frac{1}{2}\{q,s\}\right) q\sqrt{p} = 0 = \left(\frac{\partial^2}{\partial s^2} + \frac{1}{2}\{q,s\}\right)\sqrt{p} \qquad (3.3.4)$$

where $s = S_0(q)$. Equations (3.3.4) are 'equations of motion' in so far as they can be used for solving the dynamical problem (this requires writing $q\sqrt{p}$ and $\sqrt{p}$ as linear combinations of linearly independent solutions of the equation, extracting q as a function of s and inverting it to get S_0). These equations have Möbius symmetry that is signalled by the occurrence of a Schwarzian

[1]Generators of the Möbius group are translations, dilatations and 'inversions' (special conformal transformations).

[2] Faraggi and Matone 2000, section 4.2.

derivative $\{f,x\} = \frac{f'''}{f'} - \frac{3}{2}\left(\frac{f''}{f'}\right)^2$, already encountered in equation (3.2.6). Since $\{\frac{Af+B}{Cf+D},x\} = \{f,x\}$, the Schwarzian term $\frac{1}{2}\{q,s\}$ is invariant under a Möbius map[1]. This term, which can be regarded as a canonical potential, is not invariant under general coordinate transformations. This indicates that coordinate maps can be used to connect physical systems characterised by different potentials.

Equivalence Principle (EP) : For any pair of dynamical systems A and B with $W^A = V^A(q^A) - E^A(q^A)$ and $W^B = V^B(q^B) - E^B(q^B)$, there exists a coordinate transformation $q^A \rightarrow q^B$ such that $W^A(q^A)$ transforms into $W^B(q^B)$.

Since A and B are arbitrary, EP implies that any system can be mapped into that corresponding to $W^0 \equiv 0$. For the reasons given above, EP cannot be consistently implemented in classical mechanics, where S_0 satisfies (3.3.2). In that framework the W(q)'s transform as quadratic differentials: $W'(q') = (\partial_{q'}q)^2 W(q)$. As a result, W^0 can only transform into itself, and the $W(\neq 0)$ of an arbitrary system cannot be connected to W^0 by an invertible coordinate transformation. Therefore, implementing EP requires a modification of (3.3.2). Since adding a constant to S_0 changes nothing to the dynamics, the general equation that S_0 should satisfy is expected to be of the form $F(\partial_q S_0, \partial_q{}^2 S_0, \ldots) = 0$. Taking into account the fact that (3.3.2) should be the classical limit of the new equation, that equation may, with no loss of generality, be written

$$\frac{1}{2m}\left(\frac{\partial S_0}{\partial q}\right)^2 + W(q) + Q(q) = 0 \qquad (3.3.5)$$

so that the classical limit obtains as Q tends to zero. Just as in classical mechanics, all the dynamically relevant information is encoded in the functional dependence of S_0 on its argument. This solution will still be referred to as a 'reduced action', bearing in mind that it is not identical with its classical namesake.

The scalar nature of S_0 implies that W+Q transforms as a quadratic differential. On the other hand, mapping W^0 into an arbitrary $W\neq 0$ (and vice versa) is possible only if W^0 transforms with an inhomogeneous term, hence non-quadratically. In other words, under the transformation $q \rightarrow q'$, $W'(q') = (\partial_{q'}q)^2 W(q) + (q;q')$, where $(q;q')$ is the inhomogeneous term picked off

[1] This term is invariant under GL(2, **C**) (Möbius) transformations of q. A Möbius transformation of S_0 corresponds to a simple rescaling of $\frac{1}{2}\{q,s\}$.

by the transformation. It is reasonable to expect the coordinate transformation connecting W^A and W^C to be equivalent to the composition of two transformations, one mapping W^A into W^B, the other W^B into W^C. This consistency ('pseudogroup') requirement amounts to the satisfaction of the basic cocycle condition

$$(q^A;q^C) = \left(\partial_{q^C} q^B\right)^2 [(q^A;q^B) - (q^C;q^B)] \qquad (3.3.6)$$

That condition implies the Möbius invariance of $(q^A;q^B)$, which must then be proportional to the Schwarzian derivative $\{q^A;q^B\}$. In fact, (3.3.6) uniquely determines[1] the Schwarzian derivative, up to a global constant and a coboundary term. More precisely, $(q^A;q^B) = -\frac{\xi^2}{4m}\{q^A,q^B\}$.

The inhomogeneous term $(q^A;q^B)$ vanishes in the $\xi \to 0$ limit, and so does Q for a fixed W, whereby (3.3.2) is recovered with S_0 the classical reduced action.

Remarkably[2], $W(q) = (q^0;q)$, hence $W(q) = -\frac{\xi^2}{4m}\{q^0,q\}$. Such dependence of an arbitrary W on q^0 is an indication that departure from classicality relates to reconsidering the 'privileged' status of the rest frame. Since $(\partial_q S_0)^2$, which transforms as a quadratic differential under the coordinate maps, can also be expressed as the difference of two Schwarzian derivatives:

$$(\partial_q S_0)^2 = \frac{\xi^2}{2}\left(\{e^{\frac{2i}{\xi}S_0},q\} - \{S_0,q\}\right),$$

and $q^0 = \frac{Ae^{\frac{2i}{\xi}S_0} + B}{Ce^{\frac{2i}{\xi}S_0} + D}$ (where Möbius symmetry is manifest), the final expressions for W and Q are

$$W(q) = -\frac{\xi^2}{4m}\{e^{\frac{2i}{\xi}S_0},q\} \quad \text{and} \quad Q(q) = \frac{\xi^2}{4m}\{S_0,q\}.$$

Thus, consistent implementation of the Equivalence Principle leads to the equation:

$$\frac{1}{2m}\left(\frac{\partial S_0}{\partial q}\right)^2 + V(q) - E + \frac{\xi^2}{4m}\{S_0,q\} = 0\,. \qquad (3.3.7)$$

This equation is formally identical to (3.2.6) provided ξ is identified with Planck's 'reduced' constant $\hbar$. Substitution then leads to the *quantum stationary Hamilton-Jacobi equation* (QSHJE):

---

[1] Faraggi and Matone 2000, section 9.2.
[2] *Ibid.*, section 8.1.

$$\frac{1}{2m}\left(\frac{\partial S_0}{\partial q}\right)^2 + W + \frac{\hbar^2}{4m}\{S_0, q\} = 0 \qquad (3.3.8)$$

S_0 is a solution of (3.3.7) just in case $e^{\frac{2i}{\xi}S_0} = \frac{\tilde{\psi}}{\psi}$, where ψ and $\tilde{\psi}$ are linearly independent solutions of the equation

$$\left(-\frac{\xi^2}{2m}\frac{\partial^2}{\partial q^2} + V(q)\right)\psi = E\psi . \qquad (3.3.9)$$

Identifying ξ with $\hbar$ now yields the Schrödinger stationary equation. Most significantly, this equation emerges as a formal by-product of the both mathematically and physically cogent derivation of the QSHJE (3.3.8). It is the latter, rather than Schrödinger's, which directly reflects the satisfaction of the Equivalence Principle.

The 'quantum potential' $Q = \frac{\hbar^2}{4m}\{S_0, q\}$ is fundamentally different from Bohm's in that, consistently with the Equivalence Principle, S_0 can never be constant. This fact is reflected in the form

$$\psi = \left(\frac{\partial S_0}{\partial q}\right)^{-\frac{1}{2}}\left(Ae^{\frac{i}{\hbar}S_0} + Be^{-\frac{i}{\hbar}S_0}\right)$$

of a general solution to Schrödinger's equation (3.3.9), with $\xi = \hbar$. Momentum is evaluated through $p = \partial_q S_0$ ($\neq m\dot{q}$) as a function of the initial conditions, whereby a definite phase space trajectory follows. There is no need to introduce any notion of pilot wave or guidance condition.

Solving the QSHJE is facilitated by using two real and linearly independent solutions ψ and $\tilde{\psi}$ of the stationary Schrödinger equation – *this is in fact the main purpose of introducing that equation.* The appearance of both ψ and $\tilde{\psi}$, which is an indirect reflection of the underlying Legendre duality, links to the second-order linear form of (3.3.9). Since there is a many-to-one relationship[1] between reduced action and 'wave function' (a consequence of the invariance of $w = \frac{\tilde{\psi}}{\psi}$ under a Möbius transformation of $e^{\frac{2i}{\hbar}S_0}$), we are justified in regarding the Schrödinger equation as little more than a useful computational tool, which conveys less physically significant information than does the QSHJE. If it is admitted that quantum mechanics is – or rather, should be – fundamentally based upon (3.3.8), despite exclusive use of (3.3.9) since 1926, then classical mechanics is an approximation to that theory in much the same

[1] Faraggi and Matone 2000, section 14.

sense as it approximates STR in the relevant limit. Neglecting Planck's constant would then amount to relaxing the universal validity of the Equivalence Principle i.e. to restoring the privileged status of the rest frame in classical mechanics.

Writing the QSHJE in terms of the 'principal' function $S(q,t) = S_0(q) - Et$, the *time-dependent Schrödinger equation* immediately follows. The 'quantum correction' to Hamilton-Jacobi theory also retains its general structure in an Einstein-relativistic context[1] (the cocycle condition (3.3.6) remains essentially unchanged as a consistency condition on the transformation properties of W and Q). Besides that condition, which must be satisfied for (3.3.7) to follow, equivalence under coordinate transformations implies that these transformations must be locally invertible. Energy quantisation is shown to be directly related to local invertibility and the duality structure, via the imposition of an appropriate 'joining condition[2]' that *implies the existence* of a $L^2(\mathbb{R})$ (square integrable) solution of the corresponding Schrödinger equation. To the extent that the QSHJE itself is defined if and only if w is a local self-homeomorphism of the extended real line[3], and this makes it necessary for the Schrödinger equation to admit a $L^2(\mathbb{R})$ solution, energy quantisation is seen to be a direct consequence of the QSHJE, hence of the Equivalence Principle. In contrast to the usual approach, there is no need to require square integrability as an additional condition that must be satisfied by the solution of the Schrödinger equation. The fact that a particular solution is square integrable is usually taken to select the energy spectrum, determining which solutions are to be regarded as 'physical'. However, the physically relevant part of the statement is not that the solution is square integrable, but that the equation *admits* a square integrable solution. In other words, it suffices that energy values be selected as those for which such a solution exists. This does not mean, however, that for a given value of energy all solutions to the Schrödinger equation have to be square integrable (indeed, if that equation has, for a fixed E, a square integrable solution, then no other linearly independent solution can be square integrable).

Tunnel effect is reproduced without any further assumption. In particular, there is no need for invoking a probabilistic interpretation of the 'wave function' in order to account for the absence of solutions when $W(q) > 0$ for all q. Tunnel effect is found to be wholly imputable to p being a real function even in

---

[1] *Ibid.*, section 19.5.

[2] *Ibid.*, equation (15.8); see section 18.

[3] The QSHJE (3.3.8) is actually equivalent to the equation $\{w,q\} = -\frac{4mW}{\hbar^2}$, where $w = e^{\frac{2iS_0}{\hbar}}$.

Consistent implementation of the EP requires that the transformation properties of the Schwarzian derivative be satisfied. In particular, its being a Möbius invariant implies that, since under inversion ($q \to 1/q$) $0^{\pm}$ is mapped into $\pm\infty$, continuity conditions for the existence of the Schwarzian derivative must hold on the extended real line.

those regions which are classically forbidden[1] (this relates to the fact that the quantum potential[2] is never trivial).
Faraggi and Matone's EP is obviously reminiscent of its namesake in the General Theory of Relativity (GTR). So far, all approaches to 'quantum gravity' have been assuming that quantum mechanics, based upon (3.3.9) – and wherever that equation comes from – is a universal framework with which fundamental interactions have to comply, whereas gravitation is properly described in the apparently unrelated 'geometrodynamical' language of GTR. If the basic equation of (Galilean-relativistic) quantum physics is taken to be (3.3.8), there is no doubt that basic premises of the current attempts to 'quantise' gravity will have to be reconsidered.

Summarising:

1. Quantum mechanics *can* be derived from a single principle. The rationale for the introduction of a finite action quantum is 'relativistic', although in a less intuitively straightforward sense than is the case with STR.
2. The outcome of the derivation is an alternative form of the Hamilton-Jacobi equation, whose solution is a quantity having the dimensions of an action. Schrödinger's equation follows as a result of linearising that equation.
3. Energy quantisation arises as a consequence of a local invertibility condition associated with the fact that the stationary Schrödinger equation admits square integrable solutions.
4. The existence of such solutions implies that a theoretical framework that takes the Schrödinger equation (rather than the Quantum Hamilton-Jacobi equation) as its basic equation is bound by its characteristic Hilbert space structure. Alternative e.g. matrix or operator-based formulations of such a restricted formulation of quantum mechanics – which corresponds to quantum theory in its current acceptation – can be set up accordingly.

---

[1] From the CSHJE, $p = \pm\sqrt{2m(E-V)}$, implying that V – E > 0 corresponds to a forbidden region. On the other hand, from the QSHJE $p = \pm\sqrt{2m(E-V-Q)}$: because of the contribution of Q, p may be real although V – E > 0.
[2] Both classical and quantum potentials can be regarded as curvature terms in projective geometry (the Heisenberg-Born-Jordan canonical commutation relation can be expressed in terms of area functions).

5. The form of mathematical representatives of physical quantities and their mutual relationships can be derived and justified on the basis of (4) and group-theoretical considerations[1].
6. The existence of square-integrable solutions makes it possible to develop a probabilistic interpretation, in terms of the squared moduli of the relevant functions. Owing to (5), the theoretical framework is thereby turned into an effective algorithm for prediction (discussed in *Part B*). Potentially severe analytical difficulties that may be encountered while attempting to solve the Quantum Hamilton-Jacobi equation are thus bypassed. However, the cost of practical efficiency may be a loss of information about finer aspects of dynamics, which the QHJE alone affords.

Regarding (6), a parallel can be made with the ordinary probability calculus. In that framework, the 'characteristic function[2]', defined as $\Gamma(\alpha) = \int_{-\infty}^{+\infty} e^{i\alpha x} p(x)dx$, where x is a random variable and p denotes a probability distribution, finds its use in the calculation of the moments $< x^n > = \frac{1}{i^n}\left[\frac{\partial^n \Gamma}{\partial \alpha^n}\right]_{\alpha=0}$ of the distribution. Conversely, the characteristic function is determined by the moments, since $\Gamma(\alpha) = \sum_{k=0}^{\infty} \frac{(i\alpha)^k}{k!} < x^k >$. The probability distribution follows (non-uniquely in general) by a Fourier transform. As first shown by Khinchin[3] a condition, both necessary and sufficient, for a complex function Γ to be regarded as a characteristic function is that some complex-valued function φ exists, which can be normalised to unity $(\int_{-\infty}^{+\infty} \varphi^*(x)\varphi(x)dx = 1)$ and such that $\Gamma(\alpha) = \int_{-\infty}^{+\infty} \varphi^*(x)\varphi(x+\alpha)dx$. It has been suggested[4] that a square integrable solution to the Schrödinger equation might play the role of a generator of the moments of probability distributions associated with the outcomes of measuring various observables. Substituting for the classical relationship between Γ and φ the integral

$$\Gamma(\xi) = \int_{-\infty}^{+\infty} \varphi^*(x) e^{i\alpha A} \varphi(x)dx, \qquad (3.3.10)$$

---

[1] Lévy-Leblond 1974, Jordan 1975 and Section 2.2 of this thesis.
[2] Which clearly has nothing to do with the function of the same name in analytical mechanics.
[3] Khinchin 1937.
[4] Cohen 1988.

where **A** is an arbitrary Hermitian operator, basic features of the 'quantum calculus' can be retrieved – in particular, the relationship between position (x) and momentum (p) distributions automatically follows from choosing $\mathbf{A} = -i\frac{d}{dx}$ and letting $\alpha = p$. The analogy can be carried over to two and certainly more dimensions, so that formal constraints on joint distributions can be addressed. That there is no unique relationship between the probability on the one hand, and Khinchin's φ or the 'wave' function on the other suggests that the latter can give rise to as many distributions as there are variables associated with sets of compatible observables. On the other hand, it is unclear why (3.3.10) should provide the 'right' substitute in the quantum setting for the classical relationship between φ and Γ. The main virtue of parallels like this lies in their suggesting that square integrable solutions to Schrödinger's equation have just the requisite properties for setting up an effective previsional framework (assuming (5) above is satisfied).

A hint at the necessity of resorting to probability in quantum theory may be found in implications of an intriguing theorem, which has so far escaped the attention of the philosopher (as well as that of most physicists). A symplectic transformation is a mapping which, for a choice of real parameter t, takes the phase space representative M_0 of an 'initial state' into another point M_t that is a representative of the outcome of a lawful dynamical evolution (i.e. satisfying Hamilton's equations) between times t=0 and t. Symplectic transformations associated with different choices of system and different values of the parameter t form a group, and symplectic geometry concerns itself with the study of such transformations and their invariants. Liouville's theorem states that symplectic equivalence between two sets of phase space points implies equality of the phase space volumes. Equality of volumes is, however, not sufficient to guarantee symplectic equivalence for those spaces whose dimension is at least four. In 1986, another invariant, known as 'symplectic width', was discovered by Mikhail Gromov[1]. Although a symplectic transformation might appear to allow the embedding of any phase space sphere within a cylinder of the same radius (the volume of the sphere is after all finite whereas that of the cylinder is infinite), the embedding is in fact possible only if the symplectic width of the sphere is actually smaller than that of the cylinder. Since the symplectic width of both the sphere and cylinder is equal to the square of their radius $\times 2\pi$, the embeddability condition is, in that example, tantamount to requiring the radius of the sphere to be smaller than that of the cylinder.

[1] Gromov 1987; see Hofer 1998 for a survey. I thank Ivar Ekeland for drawing my attention to those references.

In mechanics, any empirical evaluation of the kinematical quantities q_i, p_i which determine a dynamical state is made with definite resolutions Δq_i and Δp_i for each degree of freedom *i*. For a given degree of freedom, the magnitude of the uncertainty is given by the 'area' $A_i = \Delta q_i \, \Delta p_i$. The smallest value η of A_i as *i* ranges over the *n* degrees of freedom then yields the minimum uncertainty or best resolution afforded by the measurement. Conservation of the symplectic width implies that, if $A_i \geq \eta$ for all *i* at the time T of the measurement, no evolution of the dynamical system from that time on can make any A_i lower than η. If there is a *k* such that $A_k \geq \varepsilon \geq \eta$ at T, then it is possible for A_k to be subsequently lower than ε, provided some A_j ($j \neq k$) at T exists such that $A_j < \varepsilon$. However, if η is a lower bound for all of the A_i, no improvement on the resolution of measurement below η is possible in principle, even if the resolution of the q_i can be improved upon at the expense of that of the p_i (and vice versa). The appearance of any such trade-off, implied by Gromov's theorem, in the framework of classical mechanics may come as a surprise. Of course, the theorem imposes no limitation on the possible a priori resolution of measurements in classical mechanics: if η is assumed to be reducible in principle to zero, then all A_i can be so reduced. In contrast, Planck's quantum of action, which in the FM scheme outlined above arises as a structural constant, may be regarded as implying an irreducible, structurally enforced lower bound on all A_i products. In fact, the authors show[1] that their equivalence principle implies phase space non-localisation, as a direct consequence of the fact that one point cannot be diffeomorphic to a line. It would be most interesting to investigate which connection the structural exclusion of phase space points as representatives of states under the equivalence principle bears with the Heisenberg inequalities, and whether it might not provide, along with the above consequences of Gromov's theorem, the key to all basic 'quantum uncertainties'. If none of the degrees of freedom of a quantum-mechanical system can possibly act as a source of accuracy and improve on the minimal resolution quantified by $\hbar$, it might well make it inevitable to resort to a probabilistic formalism, which is fortunately supplied via the square-integrability of solutions to (3.3.9).

End of Part A

[1] Faraggi and Matone 2000.

Part B

4

Quantum rules and objectification

4.1 Quantum theory, systems and preparations

In the whole of Part B, and in that part only, it will be convenient to make a distinction between *quantum theory* and *quantum mechanics*. The latter will, as in Part A, denote the theoretical framework that emerged in 1925-1926, bringing an end to the rather confused state of affairs that followed Planck's introduction of the quantum hypothesis. On the other hand, the expression *quantum theory* will be restricted to the set of elementary statements listed in Section 4.2, or some generalisation thereof. Together, these statements supply the basic ingredients and rules of a rather elementary but powerful statistical algorithm. The main purpose of this *Statistical Algorithm of Quantum Mechanics* (SAQM), which we shall take to be synonymous to 'quantum theory', is predictive: its purpose is to calculate the probabilities of the outcomes of measurements performed in specified conditions.

The elementary statements or basic rules of the SAQM are usually referred to *physical systems*. For our purposes, it will be sufficient to think in terms of *types* of (micro)physical systems, e.g. a hydrogen atom, an electron, a neutrino or a photon. Every such type can be characterised by supplying a unique list of invariant quantities such as rest mass, electric charge and a host of quantum numbers like spin, baryon number, strangeness etc. These quantities, which characterise species rather than individuals, are regarded as primitives in quantum theory. As to the heavily loaded word *state*, it will be avoided unless its relevance, use and meaning are the very subject matter of discussion. Underwriting common usage would be inappropriate since our aim in Part B is to investigate the possibility to reconstruct the whole of quantum theory with a bare minimum of presuppositions.

The word *preparation* will designate, very broadly, those conditions, including a reference to system type, that are prerequisites for the computation of predicted outcomes of measurements using the rules of quantum theory. With an abuse of language, the word will be used regardless of whether the 'preparation' in question is the result of some carefully devised experimental procedure, is just inferred from the knowledge of some prior state of affairs or is merely stated with a heuristic aim.

It may be useful to outline a typical procedure whereby a preparation is consciously achieved. The procedure basically involves (i) an appropriate source **S**, and (ii) the selection from the output of **S** of a definite (range of) value(s) of a given physical quantity. Such a selection should not be taken to imply that properties inherent to individual systems – 'their' momentum, 'their' spin (in some fixed direction) etc. – are thereby made available for further tests. As we shall see in more detail in Sections 4.3-4.4, the SAQM conflicts with simple-minded assignments of attributes to physical systems, at least as long as the attributes in question are deemed associated with the eigenvalues of the operators that represent the corresponding quantity.

We shall suppose that **S** emits electromagnetic radiation at such a low intensity that a single quantum of energy (photon) is emitted in a given time interval. The output of **S** is directed toward a polarising filter $\mathbf{P}_1$, the optical axis of which is (xx'), followed by another polariser $\mathbf{P}_2$, whose optical axis is chosen either parallel or perpendicular to (xx'). Radiation is always detected after $\mathbf{P}_2$ when the axes of the polarisers are parallel, whereas no detection ever occurs when the axes are perpendicular. It will be said that setting up the first polariser amounts to a preparation whereby the quantity "linear polarisation along the (xx') axis" can be ascertained. That preparation can now be used as a basis – in the ordinary as well as the linear algebraic sense of the word – for predicting the outcome of similar tests to be performed when $\mathbf{P}_2$ is replaced by another polariser $\mathbf{P}'$, the optical axis of which is at a known angle θ away from (xx'). One's trust in the reliability of the source and testing devices $\mathbf{P}_1$ and $\mathbf{P}_2$ is commonly expressed in counterfactual terms: whatever the outcome of detection using $\mathbf{P}'$, if detection had been attempted using $\mathbf{P}_2$ instead, the test *would* have given the expected result for that polariser.

The randomness of events like 'clicks' of a Geiger counter, impacts on a photosensitive plate, or measurement outcomes in general is an empirical fact all discussions and interpretations of quantum theory must reckon with. Because of the random character of typical occurrences, quantum-theoretical predictions of measurement outcomes are made in terms of *probability*. The probability calculus provides well-tried methods for reasoning consistently in the face of uncertainty[1]. It is also[2] in accordance with experimental practice to understand probabilities, qua objective chances, in terms of limiting frequencies in ideally infinite repetitions of a given experiment (chances, regarded as long-run frequencies, satisfy the probability axioms[3]). If, for

---

[1] Howson and Urbach 1993 ; Howson 2000.

[2] Notwithstanding the resistance of some (Fuchs 2001), a thoroughly epistemic notion of probability need not conflict with a frequency-based account of objective chances. A rather strong case can be made for the latter lending support to the proper evaluation of statistical hypotheses (Howson 2000, ch.9).

[3] These axioms, as first formulated by A.N. Kolmogorov (1933) are : (i) $p(A) \geq 0$, where A is any member of a well-circumscribed class of factual propositions; (ii) If A and B are mutually

a given preparation characterised (**P1** below) by a ket $|\Psi>$, the result a_i of measuring the physical quantity **A** obtains N_i times in N trials, the probability $p_\Psi(a_i)$, calculated using the rules of quantum theory, of that outcome conditional on that preparation is found to be in agreement with $\lim_{N\to\infty} \frac{N_i}{N}$.

4.2 Basic rules of quantum theory

> **P1**. Linear vector spaces over the field of complex numbers, equipped with an inner product (Hilbert spaces) are used as a background structure for characterising, for purposes of prevision, preparations involving physical systems of any types.
>
> **P1a**. One element of the Hilbert space **H** can be associated to any definite preparation.

This element is often called, after Dirac[1], a *ket* and more generally, albeit more questionably, a 'state vector'. Kets that differ only by a phase factor lead to the same (probabilistic) predictions, in virtue of **P6** below. Therefore, within the compass of the SAQM, such kets are to be regarded as previsionally equivalent.
Preparations for which **P1a** holds are known as *pure* cases.

> **P1a'**. There is[2] some preparation that is characterised by a given ket $|\Psi> \in$ **H**. Any such ket can be written as a linear combination or 'superposition' of kets in **H**.
>
> **P1b**. (*mixed case*) If, for any reason, no definite ket can be assigned for purposes of prevision, all of the relevant predictions may be worked out on the basis of an appropriate *statistical operator* which acts upon the elements of **H**.

A statistical operator is a self-adjoint operator whose eigenvalues are all non-negative and all add up to 1 – hence the more common name 'density operator'. A pure case statistical operator ρ is a projection operator or 'projector' ($\rho^2 = \rho$) onto the one-dimensional subspace or 'ray' in **H** spanned

---

inconsistent, then $p(A \vee B) = p(A)+p(B)$, where $\vee$ denotes the 'or' connective of propositional logic; and (iii) $p(T) = 1$, where T stands for a necessary truth. It is a consequence of the axioms that all values of p are non-negative and that they all add up to 1, suggesting that any given p(A) might be expressible as a less-than-unity ratio of appropriate quantities.

[1] Dirac 1958.

[2] It is assumed that no 'superselection rules' (Wick *et al.* 1952) forbid linear combinations of 'lawful' kets. If this happens, the projection operators that correspond to such linear combinations do not qualify as observables. Considering such possibilities, which relate to subtle group-theoretical issues, falls well beyond the scope of the present study.

by the ket which, according to **P1a**, can be ascribed to the preparation. The statistical operator assigned to a mixed case is *not* a projection operator.

> **P2**. If two preparations can be separately assigned (**P1a**) kets $|\Psi> \in \mathbf{H}$ and $|\Psi'> \in \mathbf{H'}$ respectively, then predictions bearing on any measurement to be performed jointly on Ψ and Ψ'-prepared samples can be calculated using the single ket $|\Psi,\Psi'> \in \mathbf{H} \otimes \mathbf{H'}$, where $\otimes$ denotes the tensor product of the Hilbert spaces **H** and **H'**, and $|\Psi, \Psi'>$ is an alternative way of writing $|\Psi> \otimes |\Psi'>$ (also $|\Psi> |\Psi'>$).

> **P3**. A ket $|\Psi(t_0)>$ being associated at the time t_0 to a preparation, this ket changes with time according to $|\Psi(t)> = U(t,t_0)\,|\Psi(t_0)>$, where $U(t,t_0)$ is an appropriate unitary operator[1], as long as the systems involved are not interfered with in the time interval $[t_0,t]$.
>
> **P3'**. A (pure or mixed case) statistical operator ρ evolves, in the same conditions as **P3**, from t_0 to t according to[2] $\rho(t) = U(t,t_0)\rho(t_0)U^{\dagger}(t,t_0)$.

If $\rho^2(t_0) = \rho(t_0)$ (pure case), then it follows from **P3'** that $\rho^2(t) = \rho(t)$ for $t > t_0$: a pure case cannot spontaneously evolve into a mixed case.

> **P4**. To every measurable dynamical quantity A a linear, self-adjoint[3] operator **A** can be associated, which acts upon the element of **H** defined in **P1**.

The word *observable* is commonly used to denote both a dynamical (transient) quantity and the associated operator.

> **P5**. Measuring an observable **A** on a given sample system yields a result that is uniquely associated with one of the eigenvalues a_k of the operator **A**.

Eigenvalues a_k of **A** are by definition such that $\mathbf{A}|a_{k,m}> = a_k|a_{k,m}>$ for some vector $|a_{k,m}> \in \mathbf{H}$ (that vector is not necessarily unique[4]). That the **A** operator is self-adjoint ensures that its eigenvalues, hence the results of any

---

[1] The inner (scalar) product of Hilbert space vectors provides the only Riemannian metric that is invariant under unitary transformations and is therefore preserved during the time evolution of kets.

[2] $U^{\dagger}$ is the operator adjoint to U. If an operator A admits a matrix representative with coefficients a_{mn} in some given basis, then $A^{\dagger}$ is a matrix with coefficients $a^{\dagger}_{mn} = a^{*}_{nm}$ in that basis.

[3] By abuse of language, *self-adjoint* and *Hermitian* operators will not be distinguished.

[4] For a brief discussion of such non-uniqueness or 'degeneracy', see d'Espagnat 1976 pp.17-18) and 1995 (3.1 'Remark 5').

measurement of **A**, are *real* numbers[1]. Eigenvectors $|\, a_{k,m} >$ and $|\, a_{l,n} >$ that correspond to distinct eigenvalues a_k and a_l of the same observable **A** are mutually orthogonal: their inner product $< a_{k,m} | \, a_{l,n} >$ vanishes[2]. A complete set of eigenkets of any observable defined on **H** can be used to construct an orthonormal basis on which any **H** vector can be linearly expanded. It is a distinctive feature of the SAQM that, *for any ket, there are always observables of which that ket is not an eigenvector.*

If two observables are such that the corresponding operators **A** and **B** commute (**AB** = **BA**), these two observables are said to be *compatible*: a set of eigenvectors can be found that is common to **A** and **B** (if the orthonormal basis constructed from that set is unique, then the operators are said to form a *complete set of commuting observables*). Observables whose operator representatives do not commute, i.e. such that [**A**,**B**] = **AB** – **BA** = i**C** ≠ 0, cannot have *all* of their eigenvectors in common. This implies a reciprocal relationship between the statistics of **A** and **B** measurements performed on some arbitrary preparation associated with a ket $|\, \Psi >$. The trade-off is generally expressed as an inequality[3]

$$(\Delta \mathbf{A})_\Psi (\Delta \mathbf{B})_\Psi \geq \tfrac{1}{2} |\, < \mathbf{C} >_\Psi |\,, \qquad (4.2.1)$$

where Δ**A** and Δ**B** denote the square roots of the variances or **A** and **B**. The mean value of the commutator sets a lower bound on the extent to which the dispersions given by $(\Delta \mathbf{A})_\Psi$ and $(\Delta \mathbf{B})_\Psi$ can both be reduced. It must be emphasised that (4.2.1) supposes that measurements of **A** are performed on (a large number of) elements of the prepared ensemble **E** (pure case described by the ket $|\, \Psi >$), and measurements of **B** on *other* elements of **E**. Average values $< \mathbf{A} >_\Psi$ and $< \mathbf{B} >_\Psi$ are then worked out from the data thus collected, and $(\Delta \mathbf{A})^2_\Psi = < \mathbf{A}^2 >_\Psi - < \mathbf{A} >^2_\Psi$ (ditto for **B**).

> **P6**. Let $|\, \Psi > \in \mathbf{H}$ be a *normalised* ket ($< \Psi \,|\, \Psi > = 1$) that is ascribed to a given preparation immediately prior to measuring quantity **A** on a sample system. The probability $p_\Psi(a_k)$ that the result a_k (eigenvalue of **A**) will obtain is given by $p_\Psi(a_k) = \sum_m |\, < a_{k,m} \,|\, \Psi > |^2$,

---

[1] It follows from the definition of the adjoint of an operator that $\mathbf{A} = \mathbf{A}^\dagger$ implies $a^*_{nm} = a_{mn}$, hence that all diagonal matrix elements a_{mm} are real regardless of the basis choice. The **A** matrix is diagonal in a basis of eigenkets of **A** and the corresponding eigenvalues are real.

[2] Assuming non-degeneracy, let $\mathbf{A}|\, a_i > = a_i |\, a_i >$ and $\mathbf{A}|\, a_k > = a_k |\, a_k >$; then $< a_i \,|\, \mathbf{A} \,|\, a_k > = a_k < a_i \,|\, a_k >$ and $< a_i \,|\, \mathbf{A} \,|\, a_k > = < a_k \,|\, \mathbf{A} \,|\, a_i >^* = a_i^* < a_k \,|\, a_j >^* = a_i < a_i \,|\, a_k >$, so that $(a_i - a_k) < a_i \,|\, a_k > = 0$. If $a_i \neq a_k$ then it must be the case that $< a_i \,|\, a_k > = 0$.

[3] See e.g. Merzbacher 1970, pp. 159-160 and Redhead 1987, pp. 60-61 for different proofs.

where[1] $< a_{k,m} \mid \Psi >$ is the inner product of $\mid \Psi >$ and the eigenvector $\mid a_{k,m} >$.

In the 'non-degenerate' case[2] where the subspace $\mathbf{K}^{[A,\,k]}$ spanned by the **A** eigenvectors associated with the eigenvalue a_k is one-dimensional, the probability rule simply reads

$$p_\Psi(a_k) = \mid < a_k \mid \Psi > \mid^2. \qquad (4.2.2)$$

The complex number $< a_k \mid \Psi >$ is sometimes called a probability *amplitude*. Rule (4.2.2) is often referred to as the 'Born rule' as a tribute to Max Born who first suggested it in 1926, but we shall from now on call it **PR**. A more general form of it, valid for pure as well as mixed cases, is (assuming non-degeneracy)

$$p(a_k) = \mathrm{tr}(\rho \mathbf{P}_{A,k}) = < a_k \mid \rho \mid a_k >, \qquad (4.2.3)$$

where ρ is the relevant statistical operator, $\mathbf{P}_{A,k} = \mid a_k >< a_k \mid$ the projector onto $\mathbf{K}^{[A,\,k]}$ and the trace (tr) of an operator **O** is defined as $\mathrm{tr}(\mathbf{O}) = \sum_k < v_k \mid \mathbf{O} \mid v_k >$ for any choice of orthonormal basis $\{\mid v_k >\}$ in **H**. That this trace form exists[3] is guaranteed by a famous theorem by Andrew Gleason[4]. That theorem turns out to hold only for Hilbert spaces of dimension larger than 2, but the same form is none the less adequate in two-dimensional Hilbert space. Given a statistical operator $\rho = \sum_i w_i \mid \varphi_i >< \varphi_i \mid$, where the $\mid \varphi_i >$ are not necessarily orthogonal,

$$\mathrm{tr}(\rho \mathbf{P}_{A,k}) = \sum_{m=1}^{d} \sum_i w_i < v_m \mid a_k >< a_k \mid \varphi_i >< \varphi_i \mid v_m > = \sum_i w_i \mid < a_k \mid \varphi_i > \mid^2.$$

If all the w_i but one are zero (pure case), then $\rho = \mid \varphi_i >< \varphi_i \mid$ and $\mathrm{tr}(\rho \mathbf{P}_{A,k}) = \mid < a_k \mid \varphi_i > \mid^2$. In the mixed case, where two w_i at least are nonzero, the probability is given by a sum, weighted by the w_i, where each squared modulus is the pure case expression of probability for an **A**-measurement outcome.

---

[1] The symbol $\sum_\alpha$ is to be understood as denoting either a discrete sum or an integral taken over continuously varying α's, depending on whether the eigenvalue spectrum of **A** is discrete or continuous.

[2] Degenerate spectra will be ignored in what follows, since taking degeneracy into account would make most expressions more unwieldy whilst adding nothing of conceptual interest to the issues under study.

[3] This form of the **PR** rule gives a consistent probability assignment by virtue of the projectors $\mathbf{P}_{A,k}$ being positive-semidefinite ($<\Psi \mid \mathbf{P}_{A,k} \mid \Psi > \geq 0$ for all $\mid \Psi > \in \mathbf{H}$) and adding up to the identity operator ($\sum_k \mathbf{P}_{A,k} = \mathbf{I}$).

[4] Gleason 1957 (see Hughes 1989, pp. 321-346 for a somewhat simplified proof).

For a given ket $|\Psi>$, the *expectation value* of an observable A with possible outcomes $\{a_i\}$ is

$$<\mathbf{A}>_\Psi = \sum_i a_i\, p_\Psi(a_i) = \sum_i a_i\, |<a_i|\Psi>|^2$$

$$= \sum_i <\Psi|a_i><a_i|\Psi> = <\Psi|\mathbf{A}|\Psi>,$$

using $\mathbf{A}|a_i> = a_i|a_i>$ and $<a_i|\Psi>^* = <\Psi|a_i>$. More generally, for a given statistical operator ρ,

$$<\mathbf{A}>_\rho = \mathrm{tr}(\rho\mathbf{A}). \qquad (4.2.4)$$

P7. A preparation being given, if measuring an observable **A** or any other observable **A'**, **A"**... compatible with **A** is known or expected to yield with certainty the result a_i (for **A**), a'_j (for **A'**), a''_k (for **A"**)... then the preparation can be assigned the eigenket $|u_{i,j,k...}>$ associated (in the absence of degeneracy) with the eigenvalues a_i , a'_j , a''_k ....

P7'. If such a measurement as mentioned in **P7** *has been* performed, then the statistical ensemble of those tested samples for which the result a_i (a'_j, a''_k...) has been obtained defines a new preparation characterised by the eigenket $|u_{i,j,k,...}>$ (or the statistical operator $\rho = \mathbf{P}_{A,A',...;i,j,...}$).

It should be noted that the basic rules of quantum theory do not imply that the result obtained in performing a measurement would determine the probabilities of the outcomes of measurements to be performed in the future. Rather, once any such result has been obtained, the ket assigned to the preparation is updated in accordance with **P7-P7'.**

Summing it up (in the pure case only): a Hilbert space vector is used for characterising, in a previsional rather than a descriptive sense, a definite preparation (**P1a**). To each relevant measurable quantity, there corresponds a linear and self-adjoint operator (**P4**), the real eigenvalues of which provide an exhaustive list of the possible results of measuring that quantity (**P5**). Given **P1**, the probabilities of the results of measuring some chosen quantity on a sample system are given by the probability rule **P6/PR**. Since the times of preparation and measurement do not in general coincide, the rule is generally applied after the initial vector or statistical operator has evolved according to **P3-P3'**.

4.3 The SAQM is incompatible with a simple-minded ensemble view

The above formulation of 'quantum rules' conforms to textbook tradition, but its wording carefully avoids suggesting that the probability $p_\Psi(a_k)$ calculated using **P6/PR** is that for an individual system to *be* in such or such a 'state' – a hitherto undefined notion. However, it is legitimate to ask whether $p_\Psi(a_k)$ can be interpreted as reflecting one's ignorance of definite but hitherto unknown attributes possessed, prior to measurement, by individual members of a realistically conceived ensemble. By themselves, propositions **P1(a)**, **P4**, **P5** and **P6** do not imply that a linear combination like

$$|\Psi> = \sum_i c_i |a_i>, \qquad (4.3.1)$$

where $\{|a_i>\}$ is a complete set of eigenkets of some observable **A** and more than one coefficient is non-zero, describes one ensemble, members of which would actually possess, in proportions given by $|c_i|^2$, definite attributes associated with the eigenvalues a_i of **A**. In order to find out whether this interpretation can be upheld consistently with the SAQM, let a preparation be hypothetically described using (4.3.1). It is required that $\sum_i |c_i|^2 = 1$ and the probability that the result a_k will obtain if an **A**-measurement is carried out is given by (4.2.2) i.e. $p_\Psi(a_k) = |<a_k|\Psi>|^2$. If that preparation is conceived of as an ensemble of N systems of the same type, then if the number N is large enough, it is tempting to interpret that probability as being such that a number $N_i^A \approx N p_\Psi(a_i)$ of physical systems in the ensemble possess, prior to any measurement being performed on a sample, an **A**-attribute that corresponds to the eigenvalue a_i of **A** hence, assuming no degeneracy, to the eigenket $|a_i>$. This simple reasoning is based on our experience of collections of macroscopic objects like balls in a vessel, which are believed to have such properties as size or weight regardless of and prior to any observation. Applied to a pure case (4.3.1), that simple view is found to be inconsistent with the SAQM. The conflict arises because the ket $|\Psi>$ can also be written in the basis of eigenkets $\{|b_m>\}$ of another observable **B** that does not commute with **A**. Similar reasoning as above leads to the conclusion that a number $N_j^B \approx N p_\Psi(b_j) = N |<b_j|\Psi>|^2$ of the systems in the ensemble genuinely possess, prior to any measurement, a **B**-attribute associated with the eigenvalue b_j (hence with the eigenket $|b_j>$). If so, then by analogy with balls in a vessel, one should expect the number of systems in a subensemble for which a measurement of **B** would yield b_j to be $N_j^{B'} = \sum_i N_i^A p(b_j/a_i)$, where

$p(b_j / a_i) = |< b_j | a_i >|^2$. Assuming that the probability $p(b_j)$ can be derived from a balls-in-vessel model, then

$$p(b_j) = \sum_i p(b_j / a_i)\, p(a_i) = \sum_i |< b_j | a_i >|^2 |< a_i | \Psi >|^2, \qquad (4.3.2)$$

where **PR** is applied to each term separately. There is the rub, for the basic propositions of the SAQM do not include the claim that a preparation can be broken down into any number of 'subpreparations', whereas using the **PR** rule in the last formula supposes that the calculation of $p(b_j)$ involves probabilistic contributions ($p_{a_i}(b_j) = |< b_j | a_i >|^2$) that are conditional on as many (actual or potential?) preparations as there are kets $| a_j >$, where *i* ranges over the whole spectrum of **A**. $N_j^{B'} \neq N_j^B$, belying the naive ensemble view. As is customary in quantum theory, the inconsistency occurs because of the numerical difference between the squared modulus of a sum (as in calculating N_j^B) and a sum of squared moduli ($N_j^{B'}$):

$$p_\Psi(b_j) = |< b_j | (\sum_i | a_i >< a_i |) | \Psi >|^2 = |\sum_i < b_j | a_i >< a_i | \Psi >|^2$$
$$\neq \sum_i |< b_j | a_i >|^2 |< a_i | \Psi >|^2$$

where $\sum_i | a_i >< a_i |$ is an expansion of the identity operator in **H** (closure relation for the $\{| a_i >\}$ basis). The correct prediction, i.e. the one that is in agreement with experimental data, is $p_\Psi(b_j) = = |< b_j | \Psi >|^2$ as directly calculated from (4.3.1), **PR** being applied after (implicity) summing over the amplitude products $< b_j | a_i >< a_i | \Psi >$ ($\{| a_i >\}$ can be any orthonormal basis of **H**). This is an instance of a pervasive kind of 'interference' that characterises calculations made using the SAQM.

4.4 Objectification and quantum theory

We shall assume in this section that a ket $| \Psi >$ refers to a single system **S**, designating its so-called *state*. Following Mittelstaedt[1], the hypothetical attribution to **S** of a certain (eigen)value *or* eigen*state* of an observable **A**, whether or not $| \Psi >$ is an eigenvector $| a_i >$ of **A**, is called (pure case) *objectification*. In its stronger version, the objectification hypothesis (SOH) amounts to claiming that the system actually *is in* some definite eigenstate of **A**, even if the ket ascribed to the system is (4.3.1), where more than one coefficient c_i may be non-vanishing. *Which* eigenstate it is, however, is

[1] Mittelstaedt 1998.

unknown prior to measuring **A** on **S**. According to the weaker form of the hypothesis (WOH), a definite (eigen)value a_i of **A** objectively pertains to **S** prior to its being measured. Again, that value would be subjectively unknown but no less definite. On the other hand, WOH does not assume that the 'true state' of the system is some eigenstate of **A**.

Pure case SOH entails that, regardless of the ket $|\Psi>$, an ensemble of identically prepared **S**'s amounts to a mixture of eigenstates of **A**, described by a statistical operator

$$\rho_{\Psi,A} = \sum_i w_i |a_i\rangle\langle a_i| . \qquad (4.4.1)$$

Each weight $w_i = p_\Psi(a_i) = |c_i|^2$ can be simply interpreted as a measure of our ignorance as to whether any given system in the ensemble actually *is in* the eigenstate $|a_i>$. However, we saw in the previous section that the quantum-theoretical probability $p_\Psi(b_k)$ of obtaining b_k as the outcome of measuring another observable **B**, where **[A,B]** $\neq 0$, differs from that $p_{\rho_{\Psi,A}}(b_k) = tr(\rho_{\Psi,A}|b_k\rangle\langle b_k|)$ calculated from the mixture $\rho_{\Psi,A}$ by the presence of 'interference' terms that contribute in a crucial manner to the correct evaluation of the probability. Since strong objectification would require all interference terms to vanish (or more accurately, not to appear at all) in the calculation of probabilities for *all* observables other than **A**, we must conclude that SOH is actually inconsistent with the SAQM.

Something of SOH can be salvaged by restricting its validity to **A** and to all those observables $\mathbf{C_A}$ that are compatible with it, i.e. such that $[\mathbf{A},\mathbf{C_A}]=0$. However, such a concession constitutes a severe setback if one's primary goal was to ascribe truly objective states to an individual system, for such a possibility should not be dependent upon a decision to focus on a class of quantities (and why *that* class rather than another?), with the result of excluding arbitrarily many others.

One might expect to be better off with WOH, which requires *only* that some eigenvalue of **A** pertain to **S** whatever the ket $|\Psi>$ assigned to the preparation. That ket may alternatively be written $|\Psi> = \sum_j d_j |v_j^B\rangle$, where the vectors $|v_j^B\rangle$ are eigenkets of **B**. This decomposition would appear to warrant assigning **S** some definite, merely ignored eigenvalue b_k of **B**. Letting A_i be the logical proposition "value a_i pertains to **S** given $|\Psi>$", $\neg A_i$ its negation and B_k: "value b_k pertains to **S** given $|\Psi>$", it is part and parcel of such an assignment to assume the logical equivalence of B_k and $(B_k \wedge A_i) \vee (B_k \wedge \neg A_i)$. Given the booleanity of a framework in which the equivalence normally holds, an adequate Kolmogorovian probability measure would be such that $p_\Psi(B_k) = p_\Psi(B_k \wedge A_i) + p_\Psi(B_k \wedge \neg A_i)$. Alternatively, assuming that the A_i form

a complete set of propositions exhausting the spectrum of **A**, then it is expected that $p_\Psi(B_k) = \sum_i p_\Psi(B_k \wedge A_i)$.

It might be objected that the mutual incompatibility of the two observables, expressed as the non-commutativity of **A** and **B**, precludes ascertaining the joint occurrence of a_i and b_k. The objection can be met via introducing the probability $p_\Psi(A_i \Diamond B_k)$, conditional on $|\Psi>$, that an **A**-measurement will first yield a_i, then a **B**-measurement (performed immediately afterwards) b_k. In the simplest case, using **PR** (**P6**), $p_\Psi(A_i \Diamond B_k) = |\langle a_i|\Psi\rangle|^2 |\langle b_k|a_i\rangle|^2$. Since the assignment of a ('possessed') value a_i may no longer be granted after the **B**-measurement, $p_\Psi(A_i \wedge B_k) \leq p_\Psi(A_i \Diamond B_k)$, hence $p_\Psi(B_k) \leq \sum_i p_\Psi(A_i \Diamond B_k)$. The sum $\sum_i p_\Psi(A_i \Diamond B_k)$ is equal to $tr(\rho_{\Psi,A}|b_k\rangle\langle b_k|)$, where again $\rho_{\Psi,A} = \sum_i |c_i|^2 |a_i\rangle\langle a_i|$. Therefore, $p_\Psi(B_k) \leq tr(\rho_{\Psi,A}|b_k\rangle\langle b_k|)$. Since the same inequality must clearly hold for $\neg B_k$, then[1] $p_\Psi(B_k) \geq tr(\rho_{\Psi,A}|b_k\rangle\langle b_k|)$, so that $p_\Psi(B_k) = tr(\rho_{\Psi,A}|b_k\rangle\langle b_k|)$. But this is just the probability $p_{\rho_{\Psi,A}}(b_k)$ we encountered whilst discussing SOH, leading us to the conclusion that WOH is just as inconsistent with the SAQM as SOH itself.

There again, there is a – costly – way of avoiding that conclusion. The above reasoning hinges on the additional assumption that propositions A_i and B_k combine according to the laws of classical (boolean) logic. The conclusion we have just reached could be circumvented by assuming that A_i, B_k etc. are in fact 'quantum-theoretical propositions' for which either distributivity $((B_k \wedge A_i) \vee (B_k \wedge \neg A_i)$ equivalent to $(B_k \wedge (A_i \vee \neg A_i))$ or bivalence (implying the equivalence of B_k and $B_k \wedge (A_i \vee \neg A_i)$) is not valid, *unless* **A** and **B** happen to be mutually compatible. Tailoring the propositional calculus to the projective structure of quantum theory (which connects to the non-universal compatibility of observables) must imply restrictions on the Kolmogorovity of probability, thus preventing any conflict with the SAQM. This being said, if decades of 'quantum logic' have been conclusive in any respect, it is in that no such logic offers any prospect of retaining anything of what the most lenient of realists would require of an acceptable property assignment. Besides matters of computational efficiency – which have never been an issue with quantum theory – one may also wonder whether it is advisable to introduce, be it for the sake of salvaging something of the idea of attributes being possessed by a

[1] $p_\Psi(\neg B_k) = 1 - p_\Psi(B_k)$, and the projector onto the linear subspace associated with $\neg B_k$ is the complement with respect to identity of the projector onto the subspace associated with B_k.

system, any concept of probability that does not satisfy Kolmogorov's axioms (aren't these constitutive of what is *meant* by probability?).

With respect to objectification, the situation cannot be expected to be any better in the mixed case. Two different instances of a mixture have to be distinguished. The first one corresponds to a situation where a preparation cannot be unambiguously characterised using a single ket. Rather, the predictions are based upon the assumption that the actual preparation may be any one of those associated with the (not necessarily orthogonal) vectors $|\Phi_i>$ in a given set. Each of the set members being weighted by an uncertainty measure w_i, the predictions can be derived from the statistical operator $\rho_{\Phi}^{mix} = \sum_i w_i |\Phi_i\rangle\langle\Phi_i|$ (the operator $\rho_{\Psi,A}$ above provides an example of such an ignorance-interpretable mixture). Such simple mixtures raise no additional issue over and above that of objectification for each $|\Phi_i>$ separately. Another, less straightforward instance of a mixed case arises when a statistical operator of the same ρ_{Φ}^{mix} form obtains as a result of a partial tracing operation. The question of objectification in such circumstances will be discussed in Section 7.3. The conclusion reached there is also negative: mixed case objectification is not compatible with the SAQM.

5

Deriving rules for amplitudes and 'quantum' probabilities

5.1 The composition rules for amplitudes

Simple guiding rules[1] for the calculation of probabilities can be formulated provided these are conceived of as probabilities of 'transitions'. By 'transition' is meant the connection of an 'initial' set of physical quantities (e.g. spatial co-ordinates) to a required, 'final' similar set or configuration. The initial set may be thought of as characterising a definite preparation (in the broad sense of section 4.1), whereas the final set would correspond to a given outcome of measuring a selected observable. Correct prediction requires taking into account all intermediate 'links' that may consistently connect the initial (I) and final (F) set. To each of the 'virtual paths' connecting I to F the theory assigns an *amplitude* that encodes its contribution to $I \rightarrow F$. The amplitudes combine through the application of two basic composition rules:

(α) If $I \rightarrow F$ can be broken down into subtransitions in *succession*, the resulting (total) amplitude is the *product* of the amplitudes of the successive (sub)transitions.

(β) If $I \rightarrow F$ can be conceptually analysed into *alternative*, *mutually exclusive* (virtual) paths, the resulting amplitude is the *sum* of the amplitudes for those alternative paths.

Rules α and β are similar to those that regulate the combination of *probabilities* for a Markovian process. However, the amplitudes they apply to are not identical with probabilities (they are generally complex numbers). Any one amplitude gives rise to one value of probability, computed with the standard rule

(**APR**) The probability of a transition is obtained by taking the square of the modulus of the corresponding amplitude.

APR implies that amplitudes which differ by a phase factor give rise to the same probability, and 'probability interference' is a direct consequence of the necessity of adding all the *amplitudes* assigned to alternative paths (β rule) prior to computing, using **APR**, the probability $p(I \rightarrow F)$.

---

[1] Feynman *et al.* 1965; Feynman and Hibbs 1965.

Let $I \rightarrow \{K_1, ..., K_N\} \rightarrow F$ denote N parallel sequences $I \rightarrow K_i \rightarrow F$ (i = 1 to N), where the K_i are *mutually exclusive* intermediate links in the $I \rightarrow F$ transition. Each of the sequences $I \rightarrow K_i \rightarrow F$ contributes to the total amplitude $A(I \rightarrow F)$

$$A(I \rightarrow F) = \sum_i A(I \rightarrow K_i \rightarrow F) \qquad (\beta \text{ rule}) \qquad (5.1.1)$$

where (α rule) for all i = 1 to N,

$$A(I \rightarrow K_i \rightarrow F) = A(I \rightarrow K_i)A(K_i \rightarrow F). \qquad (5.1.2)$$

The probability of $I \rightarrow F$, given by **APR**, is then

$$p(I \rightarrow F) = |\sum_i A(I \rightarrow K_i \rightarrow F)|^2 \qquad (5.1.3)$$

Connecting to the Hilbert space vector formalism is straightforward provided that the completeness of the $\{K_i\}$ is reflected in the definition of an orthonormal basis $\{|K_i>\}$. A ket associated with I is expanded in that basis as $|I> = \sum_i c_i |K_i>$, where $c_i = <K_i|I>$ and $p(I \rightarrow K_i) = |c_i|^2$ (**PR**) . The total amplitude is then

$$A(I \rightarrow F) = <F|I> = \sum_i c_i <F|K_i> = \sum_i <F|K_i><K_i|I>$$

$$= \sum_i A(I \rightarrow K_i)A(K_i \rightarrow F).$$

The rôle of the K_i is to ensure closure, as $\sum_i |K_i><K_i|$ is equivalent to the identity operator on the relevant N-dimensional Hilbert space. An alternative expression of (5.1.3) is

$$p(I \rightarrow F) = \mathrm{tr}(\rho \mathbf{P}_F), \qquad (5.1.4)$$

where $\mathbf{P}_F = |F><F|$ is the projector onto the one-dimensional subspace spanned by $|F>$ and $\rho = \mathbf{P}_I = |I><I|$.

If the experimental conditions are met for any one of the K_i to be ascertained i.e. for gathering 'which-path' information, the β rule is *not* applicable. If a measurement performed on the system *does* ascertain K_m, thereby changing its status from that of a merely 'computational' contribution to the total amplitude $A(I \rightarrow F)$ to that of an actual event, the alternative paths $I \rightarrow K_j \rightarrow F$ ($j \neq m$) have to be discarded, and the total amplitude is then

$$A(I \rightarrow F) = A(I \rightarrow K_m \rightarrow F) = A(I \rightarrow K_m)A(K_m \rightarrow F)$$

(using the α rule). In order for the β rule to cease being applicable, it is in fact enough that the K_i qualify as *experimentally ascertainable* events. Thus if perfect detectors have been set up in such a way that *one* of the K_i amongst

the N *can* be observed to be the case prior to the final configuration, each of the connecting paths $I \rightarrow K_i \rightarrow F$ then gives an individual contribution

$$p(I \rightarrow K_i \rightarrow F) = |A(I \rightarrow K_i \rightarrow F)|^2 \qquad (5.1.5)$$

to the probability $p(I \rightarrow F)$ of the $I \rightarrow F$ transition:

$$p(I \rightarrow F) = \sum_i p(I \rightarrow K_i \rightarrow F) \qquad (5.1.6)$$

The probabilities add up as they do classically i.e. in all situations where the possibility of ascertaining alternatives is tacitly granted. The probability can also be written, in contrast to (5.1.4),

$$p(I \rightarrow F) = tr(\sigma \mathbf{P}_F), \qquad (5.1.7)$$

where $\sigma = \sum_i |A(I \rightarrow K_i)|^2 |K_i><K_i|$: whereas in (5.1.4) ρ is not even a sum of K_i projectors weighted e.g. by amplitudes $A(I \rightarrow K_i)$, the weights in the mixed case statistical operator σ are the square moduli of those amplitudes (probabilistic weighting with $p(I \rightarrow K_i) = |A(I \rightarrow K_i)|^2$). In general, summing over the *amplitudes* (β) applies in those circumstances or 'experimental contexts' where no factual information can be supplied that would permit distinguishing between any two paths connecting I and F. On the other hand, the β rule expresses a requirement of completeness that must be satisfied in those situations in which the occurrence of any K_i cannot be ascertained, although an alternative choice of set-up *could* make it possible.

A sequence like $I \rightarrow \{K_1, K_2\} \rightarrow L \rightarrow F$ can be thought of as consisting of two parallel sequences $I \rightarrow K_1 \rightarrow L \rightarrow F$ and $I \rightarrow K_2 \rightarrow L \rightarrow F$, or as $I \rightarrow \{K_1, K_2\} \rightarrow L$ being followed by $L \rightarrow F$. If the probability $p(I \rightarrow F)$ is to be derived from the total amplitude, that probability should be invariant under arbitrary decompositions of the total sequence. In fact, requiring that the amplitudes associated with different decompositions of the total sequence give rise (up to a phase factor) to the same total amplitude turns out to be enough for deriving both the α and the β rule, providing that the amplitudes can be represented using analytic functions with partial amplitudes as variables.

Considering the chain $I \rightarrow K \rightarrow F$ and assigning the numbers x and y to the subtransitions $K \rightarrow F$ and $I \rightarrow K$ respectively, the total amplitude for $I \rightarrow K \rightarrow F$ will then be a function $A_\alpha(x,y)$. The form of A_α is, at this stage, entirely open[1]. Turning to the more complex chain $I \rightarrow K \rightarrow L \rightarrow F$, we let x be the amplitude assigned to $L \rightarrow F$, y that associated with $K \rightarrow L$ and z that for $I \rightarrow K$.

---

[1] $A_\alpha(x,y) \neq A_\alpha(y,x)$: it is indeed not clear whether the RHS could be associated with any meaningful transition.

Requiring invariance of the total amplitude under a change of decomposition amounts to requiring that the composition law α be *associative*, i.e. that

$$A_\alpha(x, A_\alpha(y,z)) = A_\alpha(A_\alpha(x,y), z) \qquad (5.1.8)$$

where the LHS encodes the assignment that corresponds to the sequence "I $\rightarrow$K$\rightarrow$L, then L $\rightarrow$F" and the RHS corresponds to "I$\rightarrow$K, then K$\rightarrow$L$\rightarrow$F".
In a similar way, I $\rightarrow$ $\{K_1,K_2\}$ $\rightarrow$ F will be assigned a total amplitude that is a function $A_\beta(x,y)$, where x and y stand for the amplitudes assigned to I$\rightarrow K_1\rightarrow$F and to I$\rightarrow K_2\rightarrow$F respectively. For the sequence I$\rightarrow\{K_1,K_2,K_3\}\rightarrow$F, with the same x and y, z being the amplitude assigned to I$\rightarrow K_3\rightarrow$F, associativity is expressed in the equation

$$A_\beta(x,A_\beta(y,z)) = A_\beta(A_\beta(x,y),z) \qquad (5.1.9)$$

Finally, going back to I$\rightarrow\{K_1, K_2\}\rightarrow$L$\rightarrow$F, where x, y and z are the numbers assigned to I$\rightarrow K_1\rightarrow$L, I$\rightarrow K_2\rightarrow$L and L$\rightarrow$F respectively, requiring that the aforementioned two decompositions give rise to the same amplitude implies that the equality

$$A_\alpha(A_\beta(x,y),z) = A_\beta(A_\alpha(x,z),A_\alpha(y,z)) \qquad (5.1.10)$$

has to be satisfied, i.e. α must be distributive over β.
Independent derivations of the α and β rules, both based upon earlier work by E.T. Cox[1], have been given by Y. Tikochinsky and A. Caticha[2]. Cox's original paper was not concerned at all with quantum theory. His aim was to derive, from a minimal set of assumptions, the rules of the ordinary probability calculus as the only consistent rules for combining numbers if those numbers are to be interpreted as probabilities. Cox's consistency requirements are expressed in the form of functional equations, similar to (5.1.8), (5.1.9) and (5.1.10), involving quantities that directly relate to probability. Cox showed that the basic composition laws for probabilities could be derived on the basis of those requirements alone. If a derivation of the α and β rules could be similarly achieved, such a derivation would certainly be quite significant. It would, if anything, strongly suggest that the nature and behaviour of (micro)physical objects play no essential role in constraining basic features of the SAQM. This would in turn raise the question whether most of the conceptual problems associated with 'state'-based views of quantum theory do not originate in a basic misunderstanding of the aim and structure of the theory.
It is easily seen that the functional equations (5.1.8), (5.1.9) and (5.1.10) are satisfied if one lets $A_\alpha(x,y) = xy$ and $A_\beta(x,y) = x + y$. More significantly still, the upshot of Tikochinsky's and Caticha's equivalent derivations is that these

[1] Cox 1946.
[2] Tikochinsky 1988a; Caticha 1998.

solutions also happen to be essentially the only ones. More precisely, there exists an isomorphism ϕ that maps $A_\alpha(x,y)$ and $A_\beta(x,y)$ into the product and sum of the transforms of their arguments:

$$\phi(A_\alpha(x,y)) = \phi(x).\phi(y) \qquad \text{and} \qquad \phi(A_\beta(x,y)) = \phi(x) + \phi(y)$$

Thus, through the ϕ transformation, the amplitudes combine according to the sum and product rules.

Whereas Tikochinsky's derivation is fundamentally algebraic, Caticha aims to "establish a network of relations among [experimental] setups in the hope that information about some setups might be useful in making predictions about others[1]". Caticha's work hinges on the ingenious introduction of imaginary 'filters' (essentially ideal screens with a number of holes at definite locations) and his focus is on constraints imposed on their combination, in succession or parallel. How amplitudes combine is then shown to reflect relationships among set-ups. More precisely, α-β composition expresses in a convenient and essentially unique manner the various ways complex set-ups can be consistently built up from simpler ones. It is assumed that one number may be assigned to each set-up, so that the relations among filters within set-ups translate into relations between the associated numbers (whether these numbers should be chosen complex rather than real cannot be decided). Assuming such an assignment is possible, equations (5.1.8), (5.1.9) and (5.1.10) must then be satisfied. A straightforward adaptation of Cox's method leads to the conclusion that the assigned numbers generally combine according to the sum (β) and product (α) rules, thereby justifying the designation of these numbers as 'amplitudes'.

The rules that regulate the composition of 'quantal' amplitudes are seen to be consequences of complying with three basic consistency constraints. These must be satisfied by any scheme in which numbers can be assigned, for the purpose of prevision, to a pair of events associated respectively with a preparation and an observation, taking into account all the possible links ('paths') between them. Tikochinsky takes the effectiveness of quantum theory as a motivation for the hypothesis that such an assignment is possible, whilst Caticha shows it to be implied by requiring consistency in the relationships of ideal 'set-ups'. In any case, the conclusion follows that the way amplitudes combine is not specifically 'quantal' in the sense that they would e.g. reflect the behaviour of microphysical entities. If it is excessive to claim that "quantum theory emerges as the unique way to manipulate quantum amplitudes consistently[2]", Caticha's and Tikochinsky's derivations suggest that the SAQM

[1] Caticha 2000.
[2] Caticha 1998, p. 1574.

is an instance of a general kind of predictive framework in which, as we shall see, the probabilities are computed using the **PR** rule.

5.2 From amplitudes to probabilities

It is widely believed that the (A)**PR** rule finds its sole justification in its empirical adequacy: it just happens that calculating the square of the modulus of a resultant amplitude, rather the modulus itself, its real part etc. yields the correct answer. Success in deriving the rules for the composition of amplitudes suggests, however, that deriving **PR** itself may not be out of reach. If that derivation can be successfully achieved, again, without making any reference to an underlying ontology (of particles, 'wavicles' or whatever) and any attendant notion of 'state', this would indeed provide even more incentive for re-evaluating some vexed issues (e.g. those of value-definiteness and 'quantum measurement'). Such a derivation is attempted by both Tikochinsky[1] and Caticha[2]. The latter's, which, at a purely formal level, is reminiscent of Hartle's[3], will not be discussed here. Despite his insistence on avoiding statements about the 'state' or position of a 'particle' at any given time ("These statements are not identifiable with experimental set-ups...they are not even wrong, they are meaningless[4]"), Caticha does not refrain from making concessions to more usual but potentially misleading ways of speaking: "the particle is located at x_0"; "at time t the system is in a state of wave [?] function $\Psi(t)$[5]", etc. This is especially the case in his derivation of rule **PR**, which makes an explicit reference to an ensemble of N identically prepared, independent replicas of [a] particle[6]" (in the $N\rightarrow\infty$ limit, the squared modulus rule is shown to provide the only assignment that is consistent with the law of large numbers). Even though Caticha's emphasis on set-ups and propositions suggests that his preconceptions regarding the physical nature or behaviour of physical objects involved in the experiments he discusses are minimal[7], his concessions to common usage are too pervasive in that particular issue for his

---

[1] Tikochinsky 1988b.
[2] Caticha 1998.
[3] Hartle 1968.
[4] Caticha 2000, section 5.
[5] Caticha 1998, 2000.
[6] Caticha 1998.
[7] "...by avoiding statements about the particle itself we hope to eliminate misconceptions about what the particle is and what it is actually doing between source and detector. We are not saying that the particle is either a point particle or a wave, or both, or neither. We are not saying that it went through either one hole or through another, or even that it went through both holes at the same time. In fact, beyond the fact that the particle is capable of being emitted and detected we are not assuming much at all." (Caticha 1998)

derivation of rule **PR** to fit the bill[1]. Tikochinsky's derivation, which is sketched out below, gives rise to no such qualms.

The probability $p(I\to F)$ we are looking for should take the form of a function of the amplitude $x = A(I\to F)$: $p(I\to F) = f(x)$. Tikochinsky assumes that transitions are Markovian i.e. that the probabilities of the two subtransitions $I\to K$ and $K\to F$, with amplitudes x and y, are mutually independent. If so, then

$$f(xy) = p(I\to K\to F) = p((I\to K)\&(K\to F)) = p(I\to K)p(K\to F) = f(x).f(y).$$

The amplitudes being a priori complex, the domain of the function f can be chosen as the unit disk $|x| \le 1$ (a zero-amplitude transition is regarded as one with vanishing probability: $f(0) = 0$, and $f(x) = 1$ for any complex amplitude x of unit modulus). The functional relation reduces to $f(|xy|) = f(|x|)f(|y|)$, the general continuous solution of which is

$$f(x) = |x|^k \qquad (5.2.1)$$

where $k > 0$.

The probability of the 'self-transition' $I\to I$ is, from (5.2.1), $p(I\to I) = |A(I\to I)|^k$. In order to make sense of $I\to I$ starting and finishing 'at the same time', Tikochinsky feels obliged to think of the inverse $F\to I$ of a transition $I\to F$ as proceeding backward in time. However, since transitions as we conceive them here are nothing but abstract connections established, for purposes of prevision, between two fixed configurations, there is no more need to introduce time in the picture than there was found to be any in deriving the composition rules for amplitudes. The identity of its endpoints is enough to warrant regarding the self-transition $I\to I$ as equivalent to identity, hence setting $p(I\to I)$ equal to 1. Using the α and β rules, the amplitude $A(I\to I)$ can be expressed in terms of a complete set of alternative subtransitions and their inverse: $p(I\to I) = |\sum_i A(I\to K_i)A(K_i\to I)|^k$. Letting x be the amplitude of any transition and $\bar{x}$ the amplitude of its inverse, the product and sum rules imply that $\overline{xy} = \bar{x}\bar{y}$ and $\overline{x+y} = \bar{x}+\bar{y}$. Assuming that the correspondence between x and $\bar{x}$ is one-to-one, then for any real number, $\bar{x} = x$. If the x_i are complex, the amplitude for the inverse of a transition is easily shown to be equal either to the amplitude for that transition or to the complex conjugate of that amplitude. $p(I\to I) = \left|\sum_i x_i\bar{x}_i\right|^k$ is then either equal to $\left|\sum_i x_i^2\right|^k$ or to $(\sum_i |x_i|^2)^k$, depending on whether $\bar{x} = x$ or $\bar{x} = x^*$. Choosing $\bar{x} = x^*$, then $p(I\to I) = (\sum_i |x_i|^2)^k = 1$,

[1] Caticha's derivation cannot without disfigurement be reformulated in strictly operational terms.

hence $\sum_i |x_i|^2 = 1$. Since, on the other hand, the probabilities of all the transitions $I \rightarrow K_i$ add up to one i.e. $\sum_i p(I \rightarrow K_i) = \sum_i |x_i|^k = 1$, it follows that k=2. The alternative option $\bar{x} = x$ is rejected on the grounds that there are choices of the x_i for which the sum $\sum_i x_i^2$ vanishes[1].

5.3 Destouches' derivation of the 'quantum' probability rule

Jean-Louis Destouches (1909-1980) is now a little-known figure in French post-Duhemian philosophy of science. In the early 1930s, after completing doctoral work on second quantisation, Destouches became increasingly concerned that conceptual difficulties with quantum mechanics might be deeply rooted in preconceived ideas about the aim and structure of physical theories[2]. Under the spell of Bohr's complementarity, he considered the cost of attempting to unify e.g. corpuscular and undulatory accounts of electromagnetic radiation. Unlike Bohr though, he contented such unification would require an extension of classical logic and of the rules of the ordinary probability calculus. On the other hand, in contrast to many followers of the quantum logic programme[3], Destouches did not regard any such extension as a way of coming to terms with new ontology. The key to a proper understanding of quantum theory, he contended, was to recognise that information supplied by the knowledge of the values of a given (set of) physical quantity (quantities) at a given time might not generally suffice for predicting results of measuring other quantities. He then[4] set out to develop a generic predictive scheme that would accommodate the recognised limitation.
Suppose our only concern is to evaluate the probability $p_S(Q, E_{i,Q})$ that measuring a relevant physical quantity Q on randomly chosen samples from a given preparation **S** will yield a result contained in the selected interval $E_{i,Q}$. Such probabilities are Kolmogorovian: by definition, $0 \leq p_S(Q, E_{i,Q}) \leq 1$, and for any given **S**, the probabilities of Q-measurement outcomes are completely additive: if $E_{i,Q}$ and $E_{k,Q}$ are two disjoint intervals ($E_{i,Q} \cap E_{k,Q} = \varnothing$) within which the result of measuring Q may fall, then the probability $p_S(Q, E_{i,Q} \cup E_{k,Q})$ that the result will fall in either interval is the sum $p_S(Q, E_i) + p_S(Q, E_k)$ of the probabilities that the result will fall within each interval separately. More

---

[1] For example if $x_{2p} = ix_{2p-1}$, then $|\sum_{j=1}^{2p} x_j^2|^k = 0$.

[2] Destouches 1939.
[3] E.g. Jauch 1968.
[4] Destouches 1941,1942.

generally, for any Q 'observable', $p_{\mathbf{S}}(Q, \bigcup_i E_{i,Q}) = \sum_i p_{\mathbf{S}}(Q, E_{i,Q})$ if $E_{i,Q} \cap E_{k,Q} = \varnothing$ for $i \neq k$. If σ_Q is the *spectrum* of Q i.e. the set of all the possible results that a Q-measurement may yield, or the complete set of non-overlapping intervals in which any outcome may fall, then $p_{\mathbf{S}}(Q, \sigma_Q) = 1$, since a Q-measurement will by definition yield a result that is contained in σ_Q.

Following Destouches, let us associate with any given interval $E_{i,A} \subset \sigma_A$ a set $X_0[A,E_{i,A},t_0,\mathbf{S}]$, any member $X_0[A,E_{i,A},t_0,\mathbf{S}]$ of which is called an *initial element*[1] corresponding to an '$E_{i,A}$-preparation' at t_0 i.e. to a preparation such that measuring A on any sample at t_0 is known to yield with certainty a value in $E_{i,A}$. If $E_{i,A}$ reduces to a single value $r_{i,A}$ and $X_0[A,r_{i,A},t_0,\mathbf{S}]$ is unique, then X_0 can be used to label a preparation for which a measurement of A yields the outcome $r_{i,A}$ with certainty. One further step consists in defining a function X of the time t, of t_0 and X_0 such that, if t and t_0 are fixed, a bijective relation holds between $X_0 \equiv X(t_0)$ and $X_t \equiv X(t)$. The set $X_0[A,E_{i,A},t_0,\mathbf{S}]$ is thereby mapped into another set $X_t[A,E_{i,A},t_0,\mathbf{S}]$, every member $X_t[A,E_{i,A},t_0,\mathbf{S}]$ of which we call a *predictor*[2]. If the initial conditions, i.e. the preparation or the time t_0 are changed, the predictor X_t changes with X_0. Assuming that if $X(t_i) = X(t_k)$, then $p(t_i)=p(t_k)$, X_t can be substituted for t in the expression of the probability function p: $p_X(E_{i,Q},t) \equiv p_{\mathbf{S}}(Q, E_{i,Q} ; X_t)$. Predicting measurement outcomes for any time $t > t_0$ and a given preparation **S** can now be reduced to: (i) identifying suitable initial elements X_0 at t_0; (ii) mapping these X_0 into a set of predictors X_t; and (iii) computing the probabilities of Q-measurement outcomes using those predictors. What remains to be found is the form of a probability function which, given **S** and $E_{i,Q}$, yields the value of $p_{X_t}(E_{i,Q})$.

We now suppose that the range of a quantity Q can be written as a union $E_1 \cup E_2 \cup \ldots \cup E_n$, where the intervals E_k do not overlap. For each E_i there is at least one predictor that labels a time-evolved preparation S_i such that a measurement of Q performed at t on a sample from $\mathbf{S}_i$ definitely yields a result in E_i, i.e. there is at least one $X_i \in X[Q,E_i,t,\mathbf{S}_i]$ for which $p_{S_i}(Q,E_i,X_i) = p_i(E_i) = 1$.

On the other hand, since the probability function is by assumption completely additive, we have[3] $p_X(E) = \sum_i p_X(E_i)$, where the sum is over all the disjoint subsets E_i of $E = \bigcup_i E_i$ in the chosen decomposition D. To each E_i there corresponds a unique positive number $f_i = p_X(E_i)$, and a sequence of numbers c_i, real or complex, can be generated from those f_i. Let $|c_i| = \phi(f_i)$, so that, to each f_i, there corresponds by ϕ a unique positive number. The postulated

[1] *Elément initial* (Destouches 1942).
[2] *Elément de prévision* (Destouches 1942).
[3] Dropping the explicit reference to time.

correspondence defines a function f, such that $f_i = f(c_i)$ and $0 \leq f(x) \leq 1$ for any $x \in \{c_i\}$. The function is undetermined beside its depending on X, Q and E_i, but its role as a probability function imposes that $\sum_i f_i = \sum_i f(c_i) = 1$ and each f_i must be unique for a given c_i.

Destouches now takes a crucial step: predictors $X \in X$ can be conveniently conceived as elements of a vector space **V**: if λ is a real or a complex number and $X \in \mathbf{V}$, then $\lambda X \in \mathbf{V}$; and if $X_1 \in \mathbf{V}$ and $X_2 \in \mathbf{V}$, then $X_1 + X_2 \in \mathbf{V}$. Destouches adds one further requirement: if $X \in X$ and c is a complex or real number, then $cX \in X$ only if $|c| = 1$. Each pair (c_i, X_i) can now be written as a unique vector $Y_i = c_i X_i \in \mathbf{V}$. It is therefore equivalent to consider a complete sequence of Y_i vectors or the predictor X, since the latter is completely characterised by specifying the complete set of the c_i, hence the associated X_i. Since f is still arbitrary, we are also free to require that $f(0) = 0$, letting $0.X_k = 0 \in \mathbf{V}$ where $X_k \in X$. Two vectors $c_i X_i$ and $c'_i X'_i$ will be predictively equivalent (with respect to Q, **D**) if X_i and X'_i are both members of $X[Q, E_i, t, S]$ and $f(c_i) = f(c'_i) = f_i$. We may thus write $|c_i| = |c'_i| = \phi(f_i)$, hence $c'_i = c_i e^{i\theta}$ for some arbitrary θ.

The whole purpose of Destouches's introduction of vector space structure is to write any predictor as a vector sum over the complete set of the Y_i i.e. as a linear combination of the corresponding X_i with coefficients ('amplitudes') c_i : $X = \sum_i c_i X_i$.

Now, $p_X(E) = p_X(\bigcup_i E_i) = \sum_i p_X(E_i) = \sum_i p_X(E_i) p_i(E_i)$ since[1] $p_k(E_i) = \delta_{ik}$.

Therefore[2], $p_X(E) = \sum_i f(c_i) p_i(E)$, where each coefficient $f_i = f(c_i)$ gives the value of the probability that the result will fall in E_i .

Its probabilistic role suggests that f should take the same form for any choice of decomposition. Consider another decomposition D' into disjoint sets $E'_1, \ldots, E'_j, \ldots$, such that every E_k in **D** is a union of E'_j 's (e.g., $E_1 = E'_1 \cup E'_2$, $E_2 = E'_3 \cup E'_4 \cup E'_5$ etc.). Since **D** and D' are just distinct partitions of the same 'spectrum', probabilities like $p(E_1)$ and $p(E'_1 \cup E'_2)$ should be equal, etc. What is imposed, in other words, is a consistency constraint on the values of probability 'across' decompositions, hence on the form of f. Following Destouches, we shall consider here only the case of two decompositions **D**

---

[1]This amounts to regarding $\{X_i\}$ as an orthonormal basis of **V** (Destouches 1942, pp. 542-547).

[2]If $E_i \subset E$, then $p_i(E) = 1$, and 0 otherwise.

and D' such that any E'_i in D' is a subset[1] of just one E_k in D. Letting reference to E and Q implicit, we can write

$$p(X) = \sum_i f(c_i)p(X_i) = \sum_j f(c'_j)p(X'_j),$$

where the X'_j are predictors associated with D', $X = \sum_j c'_j X'_j$ and $f(c'_j) = p_X(E'_j)$.

Since $p(X_i)$ itself can be expanded in terms of the X'_j : $p(X_i) = \sum_j f(c''_{ij})p(X'_j)$,

then $p(X) = \sum_{ij} f(c_i)f(c''_{ij})p(X'_j)$, where the coefficients c''_{ij} are those of the expansion of X_i in the $\{X'_j\}$ basis: $X_i = \sum_j c''_{ij} X'_j$. Consistency requires that

$$f(c'_j) = \sum_i f(c_i)f(c''_{ij}). \qquad (5.3.1)$$

Because there is just one E_i such that $E'_j \subseteq E_i$, (5.3.1) reduces to

$$f(c'_j) = f(c_i)f(c''_{ij}). \qquad (5.3.2)$$

Besides, $X = \sum_{ij} c_i c''_{ij} X'_j$ and $X = \sum_j c'_j X'_j$ imply that c'_j and the sum $\sum_i c_i c''_{ij}$ must coincide.

A coefficient c_j gives rise to the same predictions as c'_j if the same probability follows, i.e. $f(c_k) = f(c'_k)$, hence $|c_k| = |c'_k|$. Since $E'_j \subseteq E_i$, there is only one *i* for which c''_{ij} is not zero. Therefore, $c'_j = c_i c''_{ij} e^{i\theta}$ and $f(c_i c''_{ij} e^{i\theta}) = f(c_i)f(c''_{ij})$. The (a priori complex) numbers c_i and c''_{ij} are not subject to any condition further than those imposed on $f(c_i)$ and $f(c''_{ij})$. Any complex numbers x and y in the range of the c_i and c''_{ij} respectively must satisfy the functional equation

$$f(xye^{i\theta}) = f(x)f(y). \qquad (5.3.3)$$

This is precisely the equation that must be satisfied, according to Tikochinsky[2], by probabilities when the latter are required to be functions of 'quantum' amplitudes. However, (5.3.3) is here obtained without making any reference to 'transitions', to considerations involving time or Markovian properties. In the present framework, the functional equation results from requiring that the calculation of probabilities be consistent under arbitrary[3] choices of decomposition of the spectrum of a measurable quantity. Rejecting constant solutions like f(x) = 0 or f(x) = 1 as trivially inadequate, and imposing the further condition that f should be continuous, then the general form of the solution, identical with (5.2.1), is

---

[1] Destouches 1942, pp. 529-537. The author claims to have shown that relaxing this assumption makes no difference to the final form of the function f. However, to the best of my knowledge this more general derivation has never been published.

[2] Tikochinsky 1988b.

[3] See, however, note 3 p. 93.

$$f(x) = |x|^k \qquad (5.3.4)$$

where the constant k is a positive number[1]. This is the most general form of a continuous function f used for calculating the probabilities of Q-measurement outcomes given prior knowledge of the preparation on which those measurements are to be made. The predictive programme involves the selection of an appropriate set of X_i; writing the predictor X evolved from a given initial element X_0 as a linear combination of those X_i, where each coefficient c_i is such that the probability $p(Q, E_{i,Q}, X)$ is equal to $f(c_i) = |c_i|^k$ and k is an as yet unspecified positive number.

We are but a small step away from rule **PR**. Destouches, however, leaves it at that, merely suggesting that, since the SAQM provides a sufficiently clear instance of the kind of linear framework for prevision outlined above, the value 2 should be ascribed to k for a certain class of theories to which the quantum theory belongs[2]. This class would comprise all those theories in which no 'state magnitude' (*grandeur d'état*) exists, from which all other physical quantities of interest can be derived. Unlike, e.g., classical mechanics, in which dynamical variables are all functions of generalised coordinates or their derivatives, such theories imply necessary trade-offs associated, in quantum theory, with the non-commutation of 'incompatible' observables. Destouches's arguments in favour of the distinction are dated both in their content and expression. Nevertheless, that k *must* be equal to 2 was subsequently argued by Paulette Destouches-Février[3] as part of her own reflections[4], initiated by her husband, on the nature and aim of physical theories. Making sense of her logico-algebraic approach to the issue calls for a reminder of its background.

In an influential paper[5], Garrett Birkhoff and John von Neumann (1936) set out to examine to what extent elementary propositions stating observational outcomes, and operationally meaningful combinations of such propositions connect to relevant structures of Hilbert space. In classical analytical mechanics, a set of phase space coordinates defines a 'state', from which all of the quantities of interest to the characterisation of the motion can be derived. In order for phase space to be "imbued with reality[6]", it should be possible to correlate any of its various subsets to what Birkhoff and von Neumann call *experimental propositions*. Such a proposition is

---

[1] If discontinuities are permitted, then $f(x) = |x|^k \xi(x)$, where $\xi(x)$ reduces to 1 in the completely continuous case (Destouches 1942, pp. 534-536).

[2] Destouches 1942, pp. 541-542.

[3] Février 1951 (an earlier sketch of the argument appeared in Février 1946).

[4] Février 1951, 1956.

[5] Birkhoff and von Neumann 1936 - reprinted in Hooker 1975, vol.1 (page numbers refer to that volume).

[6] *Ibid.*, p.3.

a statement to the effect that the outcome of measuring some quantity lies in a definite subset of an adequately defined "observation space[1]". The unrestricted compatibility of observable quantities in pre-quantum physics always made it possible to establish an unambiguous correspondence between phase space subsets and definite experimental propositions. Since every boolean algebra is isomorphic to a field of sets, i.e. a non-empty set of subsets, closed with respect to complementation, union and intersection, the 'logic of classical mechanics' is boolean. As a result, there can arise no semantic difficulties with the assignment of truth values to experimental propositions, nor is there any restriction to regarding the truth of any such proposition as reflecting the possession by an individual system of a corresponding physical attribute.

In quantum mechanics, the only readily available substitute for classical phase space appears to be the Hilbert space of state vectors or so-called 'wave functions'. If so, it sounds reasonable enough to regard a closed linear subspace Σ_P of Hilbert space as a "mathematical representative" of a subset S_P of some appropriate observation space corresponding to a given set of compatible observables. An obvious candidate for the role of representative of the set-complement of S_P is the orthogonal complement ${}^{\perp}\Sigma_P$ of Σ_P. Birkhoff and von Neumann make the additional assumption that the set-theoretical product of any two representatives of experimental propositions is a valid representative of an experimental proposition[2]. This is enough to generate a propositional calculus that reflects the basic features of quantum theory.

Relations of implication between propositions are expected to correspond to set-theoretical inclusion relationships between corresponding linear subspaces of Hilbert space. Asserting that two experimental propositions P and Q are such that P implies Q (the probability of P cannot therefore exceed the probability of Q) is equivalent to asserting that the representatives of P and Q satisfy $\Sigma_P \subseteq \Sigma_Q$.In general however, as the authors point out, a closed linear subspace of Hilbert space corresponds, not to a single proposition, but to an *equivalence class* of experimental propositions. In any case, since the properties of logical implication are isomorphic to those of set-inclusion, equivalence classes of experimental propositions can in principle be correlated with subspaces of Hilbert space, and those classes then form a partially ordered system, a.k.a. a *lattice*. With complementation as introduced above, the partially ordered system is an *orthocomplemented lattice*[3].

---

[1] *Ibid.*, p.2.

[2] This automatically follows if it is conjectured that all Hermitian operators qualify as observables.

[3]Postulating that the set-product of any two representative subspaces is itself a valid representative of a class of propositions implies no restriction on the structure of the lattice.

The supremum or least upper bound of two lattice elements X and Y, sometimes written $X \vee Y$, is the unique element lying above X and Y that is also below all other elements above X and Y, where 'above' and 'below' are to be understood with respect to partial ordering[1]. The infimum or greatest lower bound of X and Y, written as $X \wedge Y$, is the unique element below X and Y in the lattice that is also above all other elements which are below X and Y. The 'meet' ($\wedge$) and 'join' ($\vee$) operations turn out to satisfy[2] identities which in ordinary logic are also satisfied by the AND and OR connectives. There is in fact total identification of meet and join with conjunction and disjunction if the lattice is boolean. Since the defining properties of a boolean algebra all have a set-theoretic counterpart, a propositional calculus which admits as mathematical representatives (in the sense of Birkhoff and von Neumann) *subsets* of a 'phase' space must be isomorphic to a boolean algebra. The lattice of experimental propositions in classical physics is boolean, implying that the meet (join) of any two propositions is a valid proposition – a conjunction (disjunction) of those propositions. Truth values can be consistently assigned to those equivalence classes of experimental propositions that correspond to nodes in the lattice. The booleanity of the lattice implies mutual distributivity of join and meet[3] i.e. $X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$ and $X \vee (Y \wedge Z) = (X \vee Y) \wedge (X \vee Z)$.

The relative independence of 'classical' propositions, which allows various measurement readings to be logically combined into a meaningful proposition, no longer holds in the quantum setting. Rather than that of a boolean algebra, equivalence classes of experimental propositions have the structure of an orthocomplemented but *non-distributive* lattice. This arises because their 'representatives' are not simple subsets but *linear subspaces* of a vector (Hilbert) space with specific projective properties. Constraints, which relate to the non-universal compatibility of observables, are thereby forced upon the propositional calculus. The apparent breakdown – or rather, non-applicability – of $(\wedge,\vee)$-distributivity in quantum theory can be traced back to more basic limitations affecting the composition of experimental propositions.

Any finite-dimensional orthocomplemented lattice that satisfies a 'modular identity[4]' is the direct product of a finite boolean algebra and a finite number of abstract projective geometries. Conversely, the direct product of a finite number of projective geometries and a finite boolean algebra is an orthocomplemented (orthomodular) lattice. A necessary and sufficient condition for such a lattice to be equivalent to a single projective geometry

---

[1] See Hughes 1989, p. 185.

[2] Hooker 1975, p.8. For a review of the properties of lattices, see Hughes 1989, pp.186-190 or Bub 1999.

[3] A Boolean algebra is an orthocomplemented *distributive* lattice.

[4] If X, Y and Z are elements of the lattice and $X \subset Z$, then $X \vee (Y \wedge Z) = (X \vee Y) \wedge Z$.

is that it contains no element U, distinct from the minimum and maximum of the lattice, such that for all X, (U∧X)∨(U∧¬X) = U. To such a lattice element would correspond a projection operator that would commute with *all* observables[1]. This, however, would be inconsistent with certain group-theoretical conditions that must be satisfied by projection operators in quantum theory[2]. Providing dimensions are both finite and larger than 2, a propositional calculus that correctly reflects such restrictions to mutual compatibility as encountered in quantum theory must have the same structure as an abstract projective geometry. Given a (not necessarily commutative) field F a condition, both necessary and sufficient, for a $(n-1)$-dimensional projective geometry $P_{n-1}(F)$ to allow orthogonal complementation is that the field admits an involutory anti-isomorphism[3] w with a definite diagonal Hermitean form $\sum_{i=1}^{n} w(x_i)\gamma_i\xi_i$, where $\gamma_i \in F$ are fixed and such that $w(\gamma_i)=\gamma_i$, and $\sum_{i=1}^{n} w(x_i)\gamma_i x_i = 0$ implies $x_i = 0$ for all $i=1$ to n. $\xi(\xi_1, ...,\xi_n)$ and $x(x_1, ..., x_n)$ being n-uples of elements of F, the sum $\sum_{i=1}^{n} w(\xi_i)\gamma_i x_i$ can be used to define an inner product $< \xi , x >$, and the orthogonality of ξ and x then corresponds to $< \xi , x > = 0$. As long as F admits an involutory anti-isomorphism that satisfies the above conditions, the choice of field remains almost arbitrary, but the options can be reduced to three – real, complex or quaternion – by imposing a locally compact topology[4] on the projective geometry. Making the connection to the SAQM requires choosing F as the field of complex numbers, in which case $w(x) = x^*$, the complex conjugate of $x \in F$, implying that the γ_i are real (they can be chosen equal to 1 to ensure normalisation). Thus, the classical vs. quantum-theoretical opposition boils down to that of boolean vs. non-boolean algebras (lattices) or to the distinction between set-theoretical structures and projective geometries. Rather than encourage the view that logico-algebraic features, abstracted from the basic formalism of quantum mechanics, somehow infringe upon the tacit acceptance of a boolean norm, Birkhoff and von Neumann suggest that quantum theory thereby exhibits greater 'logical' coherence than does any boolean-based framework. None the less, the question remains of whether meet and joint should can be imbued with semantic content, since interpretations in terms of conjunction and disjunction are no longer applicable.

---

[1]The foregoing identity is automatically satisfied if ∧, ∨ denote the logical AND, OR connectives (∧ then distributes over ∨ and X∨¬X is a tautology).
[2] see Hooker 1975, note 28 p.26
[3] Hooker 1975, p.14.
[4] *Ibid.*, note 31 p. 26.

Far from those lofty concerns, P. Février's inference[1] to the value $k = 2$ in (5.3.4) comes down to requiring consistency between a projective structure on the one hand, and the algebra of suitable representatives (in Destouches's scheme) of experimental propositions. To any proposition $P^{[Q,i]}$ of the form "the result of measuring Q at t_0 lies in $E_{i,Q}$" there can be made to correspond – albeit not biunivocally – a closed linear manifold of the vector space of predictors. The manifold associated with a given (equivalence class of) proposition(s) is that which is generated by the $\{X_i\}$ that are attached to the proposition(s). Asserting that the result of measuring both Q and R at t_0 falls in $(E_{Q,i}\times\sigma_R)\cup(\sigma_Q\times E_{R,j})$, where σ_Q and σ_R are the 'spectra' of Q and R respectively, corresponds to the weak logical sum of the propositions $P^{[Q,i]}$ and $P^{[R,j]}$ (if Q and R are the same, the assertion amounts to saying that the result of measuring Q falls in $E_{Q,i}\cup E_{Q,j}$). Similarly, the logical product of $P^{[Q,i]}$ and $P^{[R,j]}$ is taken to correspond to measurements of both Q and R at t_0 yielding a result in $(E_{Q,i}\times\sigma_R)\cap(\sigma_Q\times E_{R,j})$. The logical product is false if, for any reason, Q and R cannot be simultaneously measured. Establishing such a correspondence between elementary propositions and linear manifolds also leads to associating the negation of a proposition with the orthocomplement of the relevant manifold. The resulting structure is that of an orthocomplemented lattice. The non-distributivity of the logical product with respect to the weak logical sum is a signal that some observables cannot be simultaneously measured: the incommensurability or incompatibility of certain physical quantities is, as it were, coded into the combinatorial properties of the relevant linear manifolds, and the underlying lattice structure should be equivalent to a single projective geometry. In particular, the fundamental properties of orthocomplementation[2] are satisfied if the associated field F admits an involutory anti-isomorphism, which can be understood as defining an inner product. These are the basic observations underlying P. Février's derivation, which is here simplified and updated in terminology.

In Destouches's formalism, the decomposition into two parts of a given predictor X is always possible: each part then corresponds respectively to the truth or to the falsity of the experimental proposition: "measuring Q at t yields a result within $E_{Q,P}$". X can be written as a linear combination $X = aX_P + bX_{\neg P}$, where X_P and $X_{\neg P}$ are vectors in V_n, whose representatives are elements of the closed linear subspaces Σ_P and $\Sigma_{\neg P} \equiv {}^{\perp}\Sigma_P$ respectively.

---

[1] Février 1951.

[2] ${}^{\perp\perp}\Sigma_P = \Sigma_P$; $\Sigma_P \cap {}^{\perp}\Sigma_P = \varnothing$; $\Sigma_P \oplus {}^{\perp}\Sigma_P = \mathbf{V}$; $(\Sigma_P \subseteq {}^{\perp}\Sigma_P) \rightarrow (\Sigma_P = \varnothing)$, where $\Sigma\oplus\Xi$ denotes the smallest closed subspace of **V** that contains both Σ and Ξ.

X, X_P and $X_{\neg P}$ can also be expanded in the $\{X_i\}$ basis as

$$X = \sum_i c_i X_i \ , \quad X_P = \sum_i a_i X_i \quad \text{and} \quad X_{\neg P} = \sum_i b_i X_i .$$

Consistency between the expressions of X in the bases $\{X_i\}$ and $\{X_P\ ,\ X_{\neg P}\}$ then requires that

$$c_i = aa_i + bb_i \text{ for all } i = 1 \text{ to } n. \qquad (5.3.5)$$

Given the form (5.3.4) of Destouches's probability rule, it must then be the case that

$$\sum_i |c_i|^k = 1 \qquad (5.3.6)$$

$$\sum_i |a_i|^k = 1 \qquad (5.3.7)$$

$$\sum_i |b_i|^k = 1 \qquad (5.3.8)$$

$$|a|^k + |b|^k = 1 \qquad (5.3.9)$$

From (5.3.5) and (5.3.6), it follows that $\sum_i |aa_i + bb_i|^k = 1$ or equivalently

$$\sum_i (|a|^2|a_i|^2 + |b|^2|b_i|^2 + a^* b a_i^* b_i + a b^* a_i b_i^*)^{\frac{k}{2}} = 1 \qquad (5.3.10)$$

Orthocomplementation implies the existence of an involutory anti-isomorphism w, i.e. of an inner product such that $X_P . X_{\neg P} = \sum_i w(a_i)\gamma_i b_i = \sum_i a_i^* b_i = 0$, where $w(a_i) = a_i^*$, F being chosen as the field of complex numbers, and $\gamma_i = 1$ (the vanishing of $X_P . X_{\neg P}$ then expresses the orthocomplementarity of Σ_P and $\Sigma_{\neg P}$). Since

$$X.X = \sum_i \left(|a|^2|a_i|^2 + |b|^2|b_i|^2 + a^* b a_i^* b_i + a b^* a_i b_i^*\right), \qquad (5.3.11)$$

letting $k = 2$ in (5.3.10) turns it into an expression of the normalisation of X. Alternatively,

$$X.X = |a|^2 \|X_P\|^2 + |b|^2 \|X_{\neg P}\|^2 = |a|^2 \sum_i |a_i|^2 + |b|^2 \sum_i |b_i|^2 \qquad (5.3.12)$$

and $\|X\|^2 = 1$ with $k = 2$ in (5.3.7) and (5.3.8) makes (5.3.12) identical to (5.3.9). Thus, predictive vectors are normalised consistently with requiring that probabilities add up to unity provided the value of k is set equal to 2.

This is essentially the same requirement from which Tikochinsky would, more than three decades later, derive the same general form of the probability function, i.e.

$$f(x) = |x|^2 \qquad (5.3.13)$$

The rule $p_X(E_{i,Q}) = |c_i|^2$ whereby the probabilities of Q-measurement outcomes are computed is thus identical to rule **PR**.

Though Destouches's framework is clearly intended to recover basic features of quantum theory (SAQM), the above derivation of **PR** is not a *petitio principi* (no more, indeed, than is Tikochinsky's). Once it is accepted that predictors, as introduced by Destouches, may be conveniently represented within a vector space framework, **PR** follows from requiring consistency between various possible expressions of the probability of a given measurement outcome. Requiring that the basic form of the probability function be invariant under various expansions of a given predictor is closely related to Caticha's requirement that a given 'connection' be predictively invariant under alternative set-up choices. The fulfilment of such consistency requirements does not depend on any assumption on the nature or properties of physical systems, whether considered individually or collectively. That rule **PR** can be derived at all from such requirements suggests that quantum theory does not so much owe the 'non-classicality' of its treatment of probability to the mysteriously 'quantal' nature of microphysical objects as to having to satisfy constraints associated with the relaxation of a major (generally tacit and unquestioned) premise of classical physics: that of unrestricted access to joint values of any two 'observable' quantities.

It might be objected to Destouches's approach that his introduction of vector space structure cannot be justified beyond his intention to recover basic features of the SAQM. However, since we are after all concerned with making sense of 'quantum rules' and their implications, we may just acknowledge that the space of solutions to Schrödinger's equation happens to have the requisite mathematical properties for setting up a linear predictive scheme *à la* Destouches. That such a scheme could be set up as early as 1926 certainly proved beneficial to the rapid development of (sub)atomic physics. On the other hand, this favourable circumstance may have hampered the conceptual appraisal of such basic aspects of the theory as superpositions or probability interference. The work reviewed in this chapter, as well as more advanced developments discussed in chapter 9, strongly suggests that such features have little, if anything, to do with the enigmatic nature and behaviour of hypothetical 'quantum objects'. Rather, these are distinctive mathematical traits of a kind of predictive scheme that is instantiated by quantum theory.

6
Quantum theory as a linear scheme for prediction

6.1 Basic features

- **Bases, observables, 'eigenpreparations'**

In accordance with Destouches's[1] approach, to each outcome $E_{k,A}$ of measuring quantity **A** we associate one predictive vector, written $|u_k^A\rangle$ using Dirac's convenient 'ket' notation[2]. Vectors $|u_i^A\rangle$ such that $p_i(E_{k,A}) = \delta_{ik}$ (with the notation of section 5.3) form an orthonormal basis ($\langle u_i^A | u_k^A \rangle = \delta_{ik}$) of the n-dimensional Hilbert space $\mathbf{H}_n$. As a result, any normalised ket $|X> \in \mathbf{H}_n$ can be expanded as

$$|X> = \sum_i c_{i,A} |u_i^A\rangle . \qquad (6.1.1)$$

From each coefficient $c_{i,A} = \langle u_i^A | X \rangle$, obtained by taking the inner product of $|X>$ and $|u_i^A\rangle$ i.e. projecting $|X>$ onto the linear subspace (ray) $\Sigma_{i,A}$ spanned by $|u_i^A\rangle$ the probability of obtaining $E_{i,A}$ for a preparation associated with the ket $|X>$ can be calculated using **PR**: $p_X(E_{i,A}) = |c_{i,A}|^2$. Since $|X> = (\sum_i |u_i^A\rangle\langle u_i^A|)|X>$, $\sum_i |u_i^A\rangle\langle u_i^A|$ is just another expression of the identity operator **I**. Letting $\mathbf{P}_{A,i} = |u_i^A\rangle\langle u_i^A|$ be the projection operator such that $\mathbf{P}_{A,i}|u> = |u>$ for any $|u> \in \Sigma_{i,A}$ and $\mathbf{P}_{A,i}|u^\perp> = 0$ for any vector $|u^\perp>$ orthogonal to that ray ($<u^\perp|u> = 0 \ \forall |u^\perp> \in {}^\perp\Sigma_{A,i}$), then $\mathbf{I} = \sum_i \mathbf{P}_{A,i}$.

All that the predictive purpose of the formalism requires us to know about **A** is the exhaustive list of the values that can obtain when a measurement of **A** is performed on X-prepared samples (pure case). In order to avoid difficulties that are essentially irrelevant to our concerns, we shall henceforth assume that every interval $E_{i,A}$ reduces to a single value a_i, which corresponds to one

[1] See Section 5.3.
[2] Dirac 1958.

normalised ket $|u_i^A\rangle$. The dimension of **H** is then equal to the number of elementary items in the list $a_1,...,a_n$ of possible outcomes.
The mean value of **A** with respect to the preparation 'encoded' at time t in $|X>$ is the average over results obtained upon many repetitions of an **A**-measurement on samples from a preparation characterised by (6.1.1). It is therefore a sum of these results weighted by the corresponding probabilities:

$$< \mathbf{A} >_X = \sum_k a_k p_X(a_k) \qquad (6.1.2)$$

where **PR** has been used.
Since $< \mathbf{A} >_X = < X | (\sum_k a_k \mathbf{P}_{A,k}) | X >$, the operator to be associated with the measurable quantity **A** can be defined as a linear combination of projectors:

$$\mathbf{A} =_{def} \sum_k a_k \mathbf{P}_{A,k} \qquad (6.1.3)$$

Because a projection operator is self-adjoint, then so is **A** from (6.1.3). This 'spectral' definition of **A** is only intended to provide a predictive handle on the corresponding physical quantity. Whatever theoretical or empirical justification is to be given for the composition of the spectrum of **A** falls completely beyond the compass of quantum theory *qua* predictive framework.
Mean values are not statistically significant unless the variance of the distribution is also specified. The standard deviation $(\Delta\mathbf{A})_X$ around the mean value $< \mathbf{A} >_X$ is by definition such that

$$(\Delta\mathbf{A})_X^2 = < \mathbf{A}^2 >_X - < \mathbf{A} >_X^2 = < (\mathbf{A} - < \mathbf{A} >_X)^2 >_X \text{ i.e.}$$

$$(\Delta\mathbf{A})_X^2 = || (\mathbf{A} - < \mathbf{A} >_X) | X > ||^2 . \qquad (6.1.4)$$

Consider a 'preparation $\tilde{X}$' for which $(\Delta\mathbf{A})_{\tilde{X}} = 0$ i.e. such that the same result obtains each time a measurement of **A** is carried out. Then from (6.1.4)

$$\mathbf{A}|\tilde{X}\rangle = < \mathbf{A} >_{\tilde{X}} |\tilde{X}\rangle \qquad (6.1.5)$$

$< \mathbf{A} >_{\tilde{X}}$ is an eigenvalue of **A** that corresponds to the eigenvector $|\tilde{X}\rangle$. Owing to (6.1.3) and since $< \mathbf{A} >_{\tilde{X}}$ must be equal to one of the values a_i, $|\tilde{X}\rangle$ must be identical, up to a meaningless constant factor, to one of the vectors in $\{|u_i^A\rangle\}$ i.e.

$$\mathbf{A}|u_k^A\rangle = a_k |u_k^A\rangle , \qquad (6.1.6)$$

where $a_k \in \{a_1,\ldots,a_n\}$. Each $\left|u_i^A\right\rangle$, which satisfies the eigenvalue equation (6.1.6) with the appropriate eigenvalue a_i, can be said to label one **A**-*eigenpreparation*, i.e. a preparation that is such that measuring **A** will yield a_i with certainty. To a complete set of eigenpreparations of **A** is symbolically attached a basis set $\{\left|u_i^A\right\rangle\}$, in terms of which the representative operator for that observable can be defined. The spectral decomposition (6.1.3) of **A** is all that is needed for the purpose of predicting outcomes of measurements, provided that the quantities involved are spectrally defined in $\mathbf{H}_n$.

If two eigenvectors $\left|u_i^A\right\rangle$ and $\left|u_k^A\right\rangle$ correspond to eigenvalues a_i and a_k respectively, then $(a_i - a_k)\left\langle u_i^A \middle| u_k^A\right\rangle = 0$. Hence if $a_i \neq a_k$, then $\left\langle u_i^A \middle| u_k^A\right\rangle = 0$: eigenvectors that are associated with distinct eigenvalues of the same observable are orthogonal. Retrospectively, this provides a justification for expanding a predictive element in the $\{\left|u_i^A\right\rangle\}$ basis, for the latter provides a clear-cut distinction between the elements associated with different possible outcomes of an **A**-measurement. However, such a correspondence can be meaningfully maintained only in the case of discrete spectra. When the basic scheme outlined above is extended to continuous observables like position or momentum, the most one can achieve in a preparation is to approximate[1] 'eigenfunctions' which are at best useful fictions. Whether they label experimentally accessible dispersion-free preparations or approximations thereof, basis vectors remain a key item in the statistical algorithm.

- **Compatibility**

Two observables are said to be *compatible* just in case any eigenpreparation of one is also an eigenpreparation of the other. If **A** and **B** are compatible, then the commutator $[\mathbf{A},\mathbf{B}] = \mathbf{AB} - \mathbf{BA} = 0$. Conversely, if $[\mathbf{A},\mathbf{B}] \neq 0$, these quantities are 'incompatible': there is at least some eigenpreparation of one that is not an eigenpreparation of the other (non-commuting operators may have some eigenvectors in common ; they cannot share a complete set thereof). In other words, there exists a discrepancy between the two observables in terms of their reference sets (see Section 6.2 below). This discrepancy must show up in the analysis of statistical data, when **A**-lists and **B**-lists of outcomes are compared. If **A** and **B** do not commute, letting

[1] Dirac 1958, p.48.

[**A**,**B**] = i**C**, then standard deviations around the mean values $< \mathbf{A} >_X$ and $< \mathbf{B} >_X$ for any preparation X must satisfy the inequality[1]

$$(\Delta\mathbf{A})_X(\Delta\mathbf{B})_X \geq \frac{1}{2}| < \mathbf{C} >_X | \qquad (6.1.7)$$

The mean $< \mathbf{C} >_X$ vanishes if X is an eigenpreparation of either **A** or **B**. In general $(\Delta\mathbf{A})_X(\Delta\mathbf{B})_X > 0$ for an arbitrary preparation X: there is a reciprocal relationship between the variances of **A** and **B**, in the sense that any **A**-eigenpreparation will be statistically 'unsharp' with respect to **B**, and vice versa. The lack of 'sharpness' will affect both observables if the preparation is not predictively optimal or proper ('eigen') for either **A** or **B**. In a predictive scheme where (6.1.7) holds, no preparation can in fact be dispersion-free for *all* observables[2]. Inequalities like (6.1.7) are characteristic of linear frameworks. Their form is familiar to engineers who work with linear time-invariant systems in control, navigation and signal processing[3], and their occurrence can be traced back to a common concern with optimality[4].
(6.1.7) is correctly interpreted only in terms of sequences of measurements of **A** and **B** made on *different* X-prepared samples. As such, it has nothing to do with the question whether a truth-value can be assigned to any joint statement regarding **A** and **B** values for an individual system.

- **Mixed cases, statistical operators and a fundamental asymmetry**

A so-called *mixed* case typically combines an expression of our ignorance and the uncertainty, embodied in a ket, regarding measurement outcomes for a given pure case. More precisely, a mixed case is a probabilistic mixture of pure cases. Suppose that we are confronted with an experimental situation, which is such that the actually observed or expected distribution of the outcomes of any measurements of relevant observables is consistent with an a priori weighting of kets $|u_i>$ by probabilities w_i. Measuring a quantity **A** then amounts to estimating the mean (average) value $< \mathbf{A} >$ of **A**.

---

[1] See Merzbacher 1970, pp.158-160 for a standard proof, and Redhead 1987 pp. 60-61 for a derivation which does not explicitly rely on the Schwarz inequality (the only condition imposed is nonnegative norm). In a relatively little-known paper presented to the Prussian Academy of Sciences, Schrödinger (1930) points out that a weaker inequality can more generally be established: $(\Delta\mathbf{A})^2_X(\Delta\mathbf{B})^2_X \geq < \mathbf{F} >^2_X + \frac{1}{4} < \mathbf{C} >^2_X$, where $< \mathbf{F} > = < \frac{1}{2} (\mathbf{AB} + \mathbf{BA}) - < \mathbf{A} >< \mathbf{B} >>$ can be interpreted as the first correlation coefficient for **A** and **B** considered as two random variables. Perina *et al.* 1994 and Schroeck 1996 address situations in which $< \mathbf{F} >$ must be taken into account.
[2] See Redhead 1987, p. 62 for a short proof.
[3] See e.g. Hamming 1988.
[4] Some of the technical problems QIT theorists have been confronted with, e.g. optimising data transfer and compression, also overlap with the general concerns of the linear-algebraic theory of majorisation (Marshall and Olkin 1979 ; Bhatia 1997).

Since

$$< \mathbf{A} > = \sum_i w_i <u_i | \mathbf{A} | u_i > = tr((\sum_i w_i |u_i\rangle\langle u_i|)\mathbf{A}),$$

let us define

$$\rho = \sum_i w_i |u_i\rangle\langle u_i| \qquad (6.1.8)$$

as a statistical or density operator for that particular mixed case. As an average of projection operators associated with all the relevant pure cases, ρ exhausts what we need to know in order to make correct predictions in the given situation ($< \mathbf{A} >_\rho = tr(\rho\mathbf{A})$) . On the other hand, the decomposition (6.1.8) of ρ is not unique in general. If degeneracy occurs (i.e. not all w_i are different), there are arbitrarily many different ways[1] of decomposing the density operator. For example, there is an infinite number of decompositions of a maximally degenerate mixed case statistical operator i.e. one for which all of the w_i are equal. Since all predictions about what can be found under any experimental investigation are based on ρ alone, all those mathematically distinct decompositions are previsionally equivalent. It is up to us to exercise our judgment and choose the representation we deem the most appropriate. While such multiplicity can be viewed as a manifestation of the inherent flexibility of a representative structure (Hilbert space), which makes it possible to change the reference set in accordance with the focus of prediction, there is quite another and rather fashionable interpretation of the mutual equivalence of various decompositions of a given ρ: the impossibility of empirically distinguishing between those decompositions would be "another manifestation of the inherent inaccessibility of quantum information[2]". Calling upon 'quantum information' might be heuristically useful. However, the relevance of that notion to discussing the multiplicity of representations of mixed cases appears to be especially questionable, for such multiplicity certainly owes as much to the weighted sum form of 'classical' averaging (meant to reflect *our* ignorance of some finer details of a preparation) as it does to the vector form of 'maximal' (pure case) predictors (kets) in a linear predictive scheme like quantum theory[3].

Recent emphasis on the 'inaccessibility' of whatever information is embodied in pure or mixed cases stems from the realisation, prompted by more mundane concerns with compressing data, optimising data transfer etc., of a fundamental asymmetry between what is required in principle to specify a pure

---

[1] Hughston *et al.* 1993.

[2] Jozsa 1998, p.62.

[3] Reading QIT literature one gets the impression that many workers in the field are convinced that the multiplicity of decompositions of density operators is a newly discovered fact (for which Hughston *et al.* 1993 are often credited), whilst it has long been well known to more philosophically-minded explorers of quantum foundations (e.g. d'Espagnat 1976).

case, and the comparatively very limited amount of information any given measurement yields. For example, specifying a given pure case in a 2-D Hilbert space requires giving two real numbers (Bloch-sphere angles) θ and ϕ (if both coefficients are real, the corresponding point lies on the equatorial line – azimuthal angle $\phi = 0$ – and the preparation is specified by one parameter θ only). The range of both numbers is continuous. On the other hand, any given outcome of a single measurement of a dichotomic e.g. spin-½ observable is consistent with an infinite number of linear combinations of eigenkets of that observable. What the outcome tells us is merely that the initial predictive ket is *not* orthogonal to the eigenket corresponding to the result or (degeneracy) that it does not belong to a subspace that is orthogonal to the one spanned by the eigenvectors associated with the actual outcome. A measurement e.g. of σ_z will yield at most 1 bit of information[1]: there are two possible outcomes, +1 or -1. Obtaining either of them is of no help to distinguish between two linear combinations of eigenkets $|z_+>$ and $|z_->$ of σ_z, e.g. $\frac{\sqrt{2}}{2}(|z_+> - |z_->)$ and $-i\frac{\sqrt{3}}{2}|z_+> + \frac{1}{2}|z_->$. Furthermore, performing any number of measurements of σ_z on identically prepared samples will not allow to distinguish between, say, $\frac{\sqrt{2}}{2}(|z_+> + i|z_->)$ and $\frac{\sqrt{2}}{2}(|z_+> - |z_->)$. The reason is simply that, if the frequencies of σ_z outcomes can be thought of as approximating the probabilities of those outcomes, the rule PR with which the probabilities are computed makes no difference between coefficients whose moduli are the same. Making the distinction requires measuring *other* spin components on sufficiently many samples prepared identically. Specifying a $\mathbf{H}_2$ ('qubit') preparation thus requires that 3 independent real numbers, e.g. the components of spin in the three directions x, y and z be known, which determines a 2×2 density matrix (the procedure is obviously inapplicable if only one system is available).

This huge asymmetry between information 'content' and information yield is just a consequence of the fact that Hilbert space vectors are not all mutually orthogonal: no single measurement can distinguish between preparations associated with mutually non-orthogonal vectors. The more overlap between the kets, i.e. the larger in modulus their inner product, the less reliable the discrimination. It will become clearer as we proceed that the 'surplus structure' of Hilbert space is required in order to accommodate the contextual aspects of the predictive scheme, which cannot be dissociated from non-commutation and non-orthogonality.

---

[1] The information yield is $\log_2 D$ bits for a D-dimensional Hilbert space, hence equal to N bits in the N-qubit case, for which $D = 2^N$.

- **Context**

An *experimental context* ($|\Psi\rangle$, **A**) is formally specified by (i) giving the representative of a preparation (ket $|\Psi\rangle$) and (ii) selecting a self-adjoint operator **A** associated with the measured observable. If $\{|a_i\rangle\}$ is a basis of eigenkets of **A**, the statistical operator

$$\rho = \sum_i w_i |a_i\rangle\langle a_i| \qquad (6.1.9)$$

where $w_i = |\langle a_i|\Psi\rangle|^2$, reproduces the probability distribution of **A** (eigen)values given a $|\Psi\rangle$ preparation, and therefore expresses our predictive position vis-à-vis **A** values for that preparation. Each $|a_i\rangle$, being an eigenket of **A**, corresponds to one definite measurable (eigen)value of **A**; the $|a_i\rangle$ are mutually orthogonal (which corresponds to the mutual exclusiveness of the corresponding eigenvalues); and none of the $|a_i\rangle$ is orthogonal to $|\Psi\rangle$ (each eigenvalue a_i associated with an $|a_i\rangle$ in (6.1.9) can be experimentally ascertained in the ($|\Psi\rangle$, **A**) context. Whereas (6.1.9) is adequate for the context ($|\Psi\rangle$, **A**) (again, it yields the same probabilities as $|\Psi\rangle$ for the outcomes of measuring A or any observable that is compatible with **A**), it is inadequate, given $|\Psi\rangle$, as a predictor for the calculation of probabilities of measurement outcomes if the 'observables' involved do not commute with **A**.

6.2 Angular and statistical distance in quantum theory

Some two decades ago, William Wootters[1] unveiled a link between the Hilbert space inner product and a statistical measure introduced by Ronald Fisher in the early 1920s[2]. By practical necessity, the number of trials of a random experiment must be finite. As a result, inevitable statistical fluctuations set a limit to the extent any two weighted coins or two 'quantum' preparations associated with different kets can be regarded as distinct on statistical grounds. Statistical distance is introduced to quantify 'how far apart' from each other any two preparations are. If the preparations are to be distinguishable at all, the absolute difference in the probabilities $p^{(i)}$ (i=1,2) each of them yields for the same occurrence (e.g. 'heads') has to be no smaller than the amplitude of a typical fluctuation. If binary alternatives only (heads vs. tails, yes/no experiment) are considered, that amplitude is given, for N trials, by the root-mean-square deviation $\Delta p^{(i)} = \left[\frac{p^{(i)}(1-p^{(i)})}{N}\right]^{\frac{1}{2}}$. A simple criterion for distinguishability (in N trials) is that the regions of uncertainty of the $p^{(i)}$ be non-overlapping: $|p^{(1)} - p^{(2)}| \geq \Delta p^{(1)} + \Delta p^{(2)}$, where each $p^{(i)}$ may represent, for example, the probability of obtaining the result 'heads' upon tossing coin No. i. Fisher's statistical distance $d(P^{(1)}, P^{(2)})$ between the two 'preparations' $P^{(1)}$, $P^{(2)}$ that consist in the selection of either coin is defined as the limit, as N tends to infinity, of $N^{\frac{1}{2}}$ times the "maximum number of probabilities intermediate" between $p^{(1)}$ and $p^{(2)}$ that may be distinguished, in the above sense, from one another in N trials. Because it is harder to distinguish between the two preparations around $p = \frac{1}{2}$ than when the probabilities are further apart, $d(P^{(1)}, P^{(2)})$ cannot simply be Euclidean distance on the probability space. Indeed, in the two-coin case, calculating $d(P^{(1)}, P^{(2)})$ yields a function of the $\cos^{-1}$ form.

Generalising to a number n of possible outcomes greater than 2 (e.g. dice throws rather than coin tosses), a preparation *i*, where $p_k \geq 0$ for all k= 1 to n and[3] $\Sigma_k p_k = 1$, can be represented by a point $M^{(i)}$ $(p_1,p_2,...,p_n)$ within a flat (n-1)-dimensional hypersurface of the probability space. For a large number N of trials, the multinomial distribution is well-approximated by a Gaussian. Two preparations $P^{(1)}$ and $P^{(2)}$ associated with points $M^{(1)}$ and $M^{(2)}$ will, as before, be called distinguishable when their uncertainty regions do not

---

[1] Wootters 1980,1981.
[2] Fisher 1922.
[3] Each $p_k^{(i)}$ is the probability of the k-labelled outcome amongst the n, given the i-th preparation.

overlap[1]. An arbitrary smooth curve, parameterised by $\lambda \in [0,1]$ and connecting distinct points $M^{(1)}$ and $M^{(2)}$ on the hypersurface, where $p(\lambda=0)$ and $p(\lambda=1)$ correspond to $M^{(1)}$ and $M^{(2)}$ respectively, has a *statistical length* that is given by

$$L = \frac{1}{2}\int_0^1 \left[\sum_{i=1}^{n} \frac{1}{p_i(\lambda)}\left(\frac{dp_i(\lambda)}{d\lambda}\right)^2\right]^{\frac{1}{2}} d\lambda \qquad (6.2.1)$$

Statistical distance is then defined as the length of the shortest curve between the two p-space points. A change of variable, letting $x_i^{(m)} = \left(p_i^{(m)}\right)^{\frac{1}{2}}$ for m=1,2, transforms the statistical length into the ordinary Euclidean length of the connecting curve in x-space:

$$L = \int_0^1 \left[\sum_i \left(\frac{dx_i}{d\lambda}\right)^2\right]^{\frac{1}{2}} d\lambda \qquad (6.2.2)$$

The normalisation of probability translates into the condition $\sum_i x_i^2 = 1$. The points must lie on the unit hypersphere centred on the origin O in the n-dimensional x-space, and the curve of minimal length (geodesic) is an arc of a great circle of that sphere. Statistical distance is[2] then just the angle between the unit vectors $\overrightarrow{OM^{(1)}}$ and $\overrightarrow{OM^{(2)}}$:

$$d(M^{(1)}, M^{(2)}) = \cos^{-1}(\overrightarrow{OM^{(1)}}, \overrightarrow{OM^{(2)}}), \qquad (6.2.3)$$

hence

$$d(P^{(1)}, P^{(2)}) = = \cos^{-1}\left(\sum_i x_i^{(1)} x_i^{(2)}\right) = \cos^{-1}\left(\sum_i \left(p_i^{(1)}\right)^{\frac{1}{2}}\left(p_i^{(2)}\right)^{\frac{1}{2}}\right) \qquad (6.2.4)$$

Although the change of variable was dictated only by the quadradic form of exponents in Gaussian approximations to the multinomial distribution, switching from the p_i to the x_i is suggestive of reasoning in terms of 'amplitudes' rather than probabilities (the relation between them being a real number version of rule **PR**).

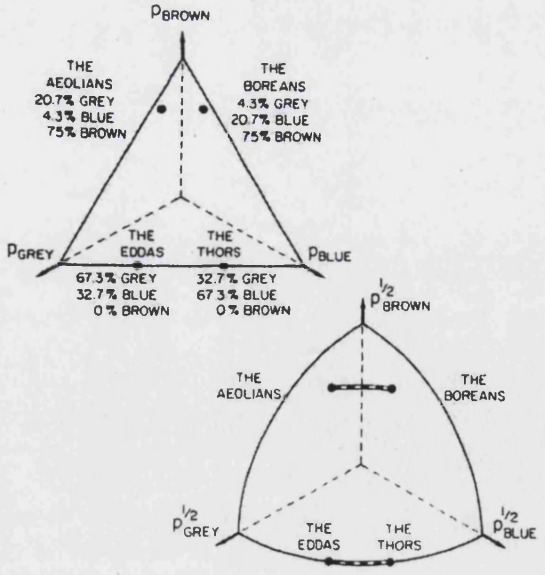
---

[1] A quantitative notion of 'uncertainty region' appropriate to the Gaussian form of the distribution is introduced by Wootters 1981, p. 360.

[2] The result also holds in the case of a countably infinite number of outcomes (Wootters 1981, p. 360).

Fisher's metric is useful to eliminate spurious distinctions that may arise as a result of the choice of a 'flat' metric[1]. This is nicely illustrated by Wheeler[2], following Wootters[3]:

> We find ourselves in the midst of a tribe of people who speak an unknown language. Are they the Eddas, who are friendly? Ore are they the Thors, who are cannibals? All we have to go on is the color of the eyes of the sixteen warriors who encircle us. Our scouts have told us that 67.3% of the Eddas have grey eyes; 32.7%, blue eyes; whereas for the Thors the proportions are the other way around. Our statisticians have told us that, if the majority of the sixteen pairs of eyes are grey, we have close to a twelve-to-one chance of being safe. And so they are – and so we are! That is distinguishability in action.
>
> Unfortunate explorers, we find ourselves on a new journey to a new continent confronted anew by the old issue. Are the sixteen who now surround us the friendly Aeolians or the deadly dangerous Boreans? At first sight, it appears that it will be much more difficult to be certain of our appraisal. Why? Because the differences are now so much less between the two tribes in count of grey and blue eyes. This conclusion bases itself (plane $p_{grey} + p_{blue} + p_{brown} = 1$ in the upper left hand diagram [below]) upon the separation of the two representative points in question in a linear probability diagram, a separation large in the one continent, small in the other.
>
> Statistical analysis, however, shows that if the grey eyes are again in the majority we again have close to a twelve-to-one assurance of being safe. The linear diagram is misleading because it is based on probabilities. To make distinguishability properly shine out, we should use not probabilities but probability amplitudes; not linearly related quantities that lie on a sector of a plane, but quadratically related quantities that lie on a sector of a sphere, $(p_{grey}^{1/2})^2 + (p_{blue}^{1/2})^2 + (p_{brown}^{1/2})^2 = 1$.
>
> In brief, the proper depiction of distinguishability demands Hilbert space. The angle in [here, real] Hilbert space between two nearly identical probability-amplitude vectors (stippled lines in the lower right-hand diagram [below])...is the proper measure of their distinguishability.



[1] Fisher himself was concerned with a statistical analysis of the phenomenon of genetic drift – Fisher 1956 (pp. 8-17); Kimura 1962.
[2] Wheeler 1988.
[3] Wootters 1980.

The 'quantum' analogues of coin or dice 'preparations' would be what Wootters calls "preparations of pure states[1]". However, one "new feature" in the quantum setting is that, "[w]hereas for dice there is only one possible experiment to perform (namely, rolling the die), for quantum systems there are many, one for each different analyzing device[2]". Such multiplicity is characteristic of those situations to which **PR** applies, its most significant aspect being that "two preparations may be more easily distinguished with one analyzing device than with another[3]". To put it differently, each type of measurement instantiates a context for the acquisition of information, and it turns out any two such contexts are not generally equivalent in that respect. On the other hand, detailed knowledge of the system-apparatus interactions required for measurement to have a definite outcome have no more role to play in the statistical analysis than they do in the case of a coin toss[4]. Characteristically, maximum distinguishability in one context implies maximal uncertainty (equal probability) in some other context, the two contexts being associated with incompatible observables.

Let $P^{(1)}$ and $P^{(2)}$ be two preparations to which a ket $|\Psi^{(1)}\rangle$ and $|\Psi^{(2)}\rangle$ respectively can be associated, **A** an observable the measurement of which is used to distinguish between the two preparations. The probability of the outcome a_k associated with the eigenket $|a_k>$ of **A** is $p^{(i)}(a_k) = |\langle a_k|\Psi^{(i)}\rangle|^2$ for preparation i=1,2.

Using **PR**, the statistical distance (6.2.4) between the two preparations in the context determined by choosing the **A** observable is

$$d(P^{(1)}, P^{(2)}) = \cos^{-1}\left(\sum_{i=1}^{N}|\langle a_i|\Psi^{(1)}\rangle||\langle a_i|\Psi^{(2)}\rangle|\right) \qquad (6.2.5)$$

If probabilities are computed from the same eigenket of **A** for either preparation, the two preparations are trivially indistinguishable and $d(P^{(1)}, P^{(2)})=0$. In contrast, if $P^{(1)}$ and $P^{(2)}$ correspond to two distinct eigenkets of **A**, they are statistically as far apart as two preparations can possibly be. Those two extremes aside, best discrimination is afforded whenever one of the preparations, say $P^{(1)}$, happens to be described by an **A**-eigenket. Because of the mutual orthogonality of **A**-eigenkets, the sum in (6.2.5) then reduces to a single term:

$$d(P^{(1)}, P^{(2)}) = \cos^{-1}(|<\Psi^{(1)}|\Psi^{(2)}>|) \qquad (6.2.6)$$

---

[1] Wootters 1981, p.360.
[2] Wootters 1981, p.361.
[3] Wootters 1981, p.361.
[4] One should thus be wary of such utterances as "with respect to a *particular* measuring *device*" (Wootters 1981, p.361 - italics mine).

with, say, $|\Psi^{(1)}> = |a_k>$ and the statistical distance between two preparations is then identical with the Hilbert space angle between the rays associated with the preparations. Thus (6.2.6) holds only provided a 'reference' observable is selected, relative to which one of the two preparations ($P^{(2)}$) is statistically 'gauged'. No such simple quantity as the angle between two Hilbert space vectors supplies a measure of statistical distance unless such a reference is provided.

Summarising:

> Given a general notion of statistical distance (6.2.4) between two preparations, then if (i) two such preparations are ascribed representatives for purposes of prediction using the resources of quantum theory, so that the probabilities of measurement outcomes are calculated using **PR**; and if (ii) a context ($\Psi^{(1)}$,**A**) is selected such that the predictive vector $|\Psi^{(1)}>$ associated with one of the preparations denotes an eigenpreparation of the observable **A**, then the statistical distance between the two preparations is the angle between the corresponding two Hilbert space rays.

Since statistical distance is determined entirely by the size of statistical fluctuations, regardless of whether the analysed preparations involve quantum mechanical systems, coins, dice or human populations, one may find it "surprising[1]" that Hilbert space angle, which is not a statistical concept, should be found to connect with it at all. How come "nature rather mysteriously makes these two kinds of distance identical[2]"? Wootters believes that Hilbert space angle separates 'quantum states', which he appears to conceive of as representations of inherent objective features of a quantum-mechanical system (the reference to nature suggests as much). Such states would be objectively 'distant' from one another, regardless of any operational procedure whereby their 'degree of likeness' could be ascertained. But then, how can "ubiquitous statistical fluctuations in the outcomes of measurements[3]", which have *a priori* nothing to do with the nature and properties of physical objects those measurements are made upon, so tightly connect to a "geometry of the set of states[4]" that is taken to have nothing *a priori* to do with statistics but everything to do with the nature and properties of physical systems? Rather than seize the opportunity to re-evaluate the cogency of referring to 'states' in relation to the SAQM, Wootters appears to seek an objective, 'natural' grounding for the equivalence: "It is as if nature defines [?] distance

[1] Wootters 1981, p.361.
[2] Wootters 1981, p.357.
[3] Wootters 1981, p.357.
[4] Wootters 1981, p. 357.

between states by counting [?] the number of distinguishable intermediate states[1]". It is little wonder that "we cannot claim to understand physically this connection between statistics and geometry[2]", if by understanding 'physically' one means getting insights into some mysterious counting (?!) that would spontaneously take place in the physical world. Should we think of Nature Herself as capable to distinguish between objective 'states', and to do so in a manner that would happen to be most convenient for the purpose of hypothetical statisticians?

Following such a track any further would be pointless. Instead, let us reformulate Wootters's sentence in a way that is more mindful of the predictive aim of the basic framework of quantum theory: "It is as if distance between pure case preparations could be defined by counting the number of operationally distinguishable[3] intermediate, virtual preparations". Rather than as one more riddle about nature, the angular-statistical distance link should be regarded, in a more positive way, as providing a statistical justification for setting up a predictive framework based on a projective, Hilbert space structure. This is the closest to an empirical grounding for a linear predictive scheme *à la* Destouches we can possibly get. The introduction (Section 5.3) of a sequence of (real or complex) numbers c_i can now be justified by regarding their moduli as expressions, given by (6.2.6), of the statistical distance between the preparation P on which the predictions are based and a complete set of 'counterfactual' eigenpreparations relative to the chosen observable **A**. Those preparations are such that every one of them, if it were realised instead of P, would yield with certainty the associated outcome a_i when an **A**-measurement is carried out. Letting $d^A_{kl} = \cos^{-1}\left(\left|\langle a_k^{(1)} | a_l^{(2)} \rangle\right|\right)$ be the statistical distance between two eigenpreparations $P_{A,k}{}^{(1)}$ and $P_{A,l}{}^{(2)}$, then $|< a_k{}^{(1)} | a_l{}^{(2)} >| = \cos(d^A_{kl}) = \delta_{kl}$. Expanding the ket $|\Psi>$ associated with P in the $\{|a_i>\}$ basis, the positive real number $|c_k|^2 = |< a_k | \Psi >|^2 = \cos^2 d(P_{A,k}, P) \leq 1$ can be regarded as the probability[4] that measuring **A** on P-prepared samples will yield a_k. The probabilities of measurement outcomes as computed using **PR** can then be understood as measures, in the unit interval, of the statistical distance between P and a complete set of counterfactual eigenpreparations that form what may be called a 'distinguishability basis'. On the other hand, Wheeler's

---

[1] Wootters 1981, p. 361.

[2] Wootters 1981, p.361.

[3] In principle, though not necessarily in practice, because of instrumental limitations that could, or could not, be overcome.

[4] In the first part of his 1981 paper, Wootters shows that the $\cos^2$ form of a probability distribution, as observed in polarisation experiments is actually the only form that is compatible with a notion of distinguishability based on the relative angle between two polariser orientations.

claims that "Wootters *derives* Hilbert space with its familiar complex probability amplitudes[1]" is incorrect, for (6.1.6) obtains only once probabilities of measurement outcomes are inserted into (6.1.5), and these are computed *using* **PR**, which supposes the very kind of structure Wheeler claims is being derived. It is none the less significant that a basic metric feature of Hilbert space should be matching statistical requirements that are essentially independent of assumptions regarding the nature and properties of physical systems.

Some QIT-related work in the 1990s has been concerned with finding ways to quantify gains of information in typical 'quantum' situations[2]. Suppose we sample once from what we know to be either of two distributions $p^{(1)}(a_i)$ or $p^{(2)}(a_i)$, a_i being a possible outcome of an **A**-test, without knowing in advance which distribution is sampled from. Prior to sampling, the most we know is the prior probabilities w_1 and w_2 of each preparation being the case. The probability of our making an error in guessing the identity of the distribution after **A**-testing a sample is defined as $p_e = \sum_k \min\{w_1 p^{(1)}(a_k), w_2 p^{(2)}(a_k)\}$. The smaller its numerical value, the better our guesswork succeeds in differentiating between the two distributions. As it stands though, p_e is unacceptable as a distinguishability measure because it leads to paradoxical results in multiple sampling[3]. On the other hand, its asymptotic form (the so-called *Chernoff bound*), which is exponential in the number N of samples tested before the decision is made, happens to be conveniently independent both of the number of trials and of the prior probabilities w_1 and w_2. Letting $p_e(N)$ be the probability of error after sampling-testing N times on the distributions, the Chernoff theorem[4] says that

$$p_e(N) \leq \left[\min_\alpha \sum_i p^{(1)}(a_i)^\alpha p^{(2)}(a_i)^{1-\alpha} \right]^N , \qquad (6.2.7)$$

where $0 < \alpha < 1$. Coming up with an analytical expression for the Chernoff bound appears to be hopeless. However, setting $\alpha = ½$ yields the simpler sum

$$\sum_i p^{(1)}(a_i)^{\frac{1}{2}} p^{(2)}(a_i)^{\frac{1}{2}} , \qquad (6.2.8)$$

which is a distinguishability measure known as *statistical overlap*. The choice of exponent is dictated by considerations of symmetry and

[1] Wheeler 1990, p.12. - Italics mine.
[2] Fuchs 1996 ; Fuchs and Peres 1996.
[3] Cover 1974.
[4] Chernoff 1952 ; Cover and Thomas 1991.

tractability. Assuming a Riemannian Fisher metric[1], the quantity $\cos^{-1}(\sum_i p^{(1)}(a_i)^{\frac{1}{2}} p^{(2)}(a_i)^{\frac{1}{2}})$, already encountered as (6.2.4), measures the extremal (geodesic) distance between the two distributions on the probability simplex. A quantum-theoretical adaptation of the same decision problem is expected to lead to a Hilbert space expression of (6.2.7), just as (6.2.5) was found to be a 'quantal' expression of (6.2.4). The problem now involves two (pure or mixed case) preparations to which the two statistical operators ρ_1, ρ_2 are associated. The observer does not know which is actually the case and assigns them the prior probabilities w_1 and w_2. The challenge is for him to perform any measurement he deems appropriate and guess the identity of the preparation on the basis of the outcome. The 'best' choice of measurement and the best guess are those that, together, minimise the expected probability of error, i.e. the probability of ρ_1 (resp. ρ_2) being the correct choice times the conditional probability that the decision based on the result of measurement will be wrong when ρ_1 (resp. ρ_2) is the 'right' preparation[2]. There again, for multiple sampling, minimising the probability of error over all possible measurements leads to an unwanted result: the optimal measurement explicitly depends on the number of measurements performed. Since working out an analogue to the Chernoff bound raises the same kind of analytical difficulties, we are again led to restrict ourselves to the symmetric case $\alpha = ½$. A rather straightforward 'quantum' version of (6.2.8) is given by

$$F(\rho_1, \rho_2) = \mathrm{tr}(\rho_2^{\frac{1}{2}} \rho_1 \rho_2^{\frac{1}{2}})^{\frac{1}{2}}. \qquad (6.2.9)$$

This quantity, often referred to as *quantum fidelity*[3], owes much of its usefulness to its being symmetrical under permutation of ρ_1 and ρ_2 and invariant under unitary transformations. If $\rho_1 = |\Psi_1><\Psi_1|$ and $\rho_2 = |\Psi_2><\Psi_2|$ (pure cases), then $F(\rho_1, \rho_2)$ reduces to the modulus of the inner product $<\Psi_1|\Psi_2>$, leading to (6.2.6). In the mixed case, where $F(\rho_1,\rho_2) \geq |<\Psi_1|\Psi_2>|$, quantum fidelity provides an upper bound for so-called *purifications*[4] of ρ_1 and ρ_2 (the 'best' such purification would give a modulus of the inner product equal to F).

---

[1] Fisher's metric plays an important role in maximum likelihood parameter estimation (Cover and Thomas 1991), where the aim is to calculate the decrease of an estimator's variance with the number of samples drawn. One of its first uses was found in estimates of the divergence between two statistical populations defined by their probability distributions (Bhattacharyya 1943).

[2] Helstrom 1976.

[3] The name comes from its providing a criterion (Jozsa 1994) for the 'fidelity' of the signals conveyed via 'quantum' channels (e.g. using photons in an optic fibre cable).

[4] Uhlmann 1976 ; Jozsa 1994.

Different approaches to the estimation problem lead to different measures of distinguishability. None of these quantities can be regarded as *the* most adequate measure: "In any case, the best measure...is the one that is relevant to the actual application in which we are interested[1]." This is one further motive for doubting that any expression of the angular-statistical link reflects objective constraints forced upon our formalisations by the natural world.

6.3 Evolving, measuring and predicting

> Between the observations nothing at all happens, only time has, 'in the interval', irreversibly progressed on the mathematical papers!
> *W. Pauli* (letter to M. Fierz, 1947)

How predictors evolve in time as long as no attempt at gathering information is made has yet to be addressed. Conservation of probability is guaranteed via introducing a unitary, hence inner-product-preserving operator **U**, such that $|X>_{t'} = \mathbf{U}(t,t')|X>_t$ for two times t and t'. The **U** operators, parameterised by the time t, must have the combinatorial properties that characterise groups: they must be invertible, their combination must be time-wise associative and there must be a neutral element **U**(t,t) that is equivalent to the identity operator on **H**. According to Stone's theorem[2], if the operators $\{\mathbf{U}_t\}$ form a weakly continuous group of unitary operators isomorphic to $(\Re,+,0)$, i.e. a group parameterised by the real numbers, then there exists a unique Hermitian operator **H** such that, for all $t \in \Re$, $\mathbf{U}_t = e^{-i\mathbf{H}t}$. If **H** is not itself time-dependent, then

$$d_t\mathbf{U}_t = -i\mathbf{H}\mathbf{U}_t \,, \qquad (6.3.1)$$

where $d_t \equiv \frac{d}{dt}$. Writing $|X>_t = \mathbf{U}_t|X>_0$, where $|X>_0$ is some fixed predictor at t=0, one gets[3]

$$i\, d_t|X>_t = \mathbf{H}|X>_t \qquad (6.3.2)$$

---

[1] Fuchs and Peres 1996.

[2] see Fano 1971 ;Hughes 1989, pp. 114-118.

[3] The burden of time-dependence may be shifted from kets to self-adjoint operators representing observables. This is the so-called 'Heisenberg picture', as opposed to the 'Schrödinger picture' of the main text. Letting $\mathbf{A}_t = \mathbf{U}_t^{\dagger}\mathbf{A}\mathbf{U}_t$, where **A** is a time-independent observable, then the time-dependent observable $\mathbf{A}_t$ satisfies the equation $id_t\mathbf{A}_t = [\mathbf{A}_t\,, \mathbf{H}]$. $\mathbf{A}_t$ is therefore time-independent iff $\mathbf{A}_t$ commutes with **H**.

Again, determining the explicit form of **H** falls beyond the compass of quantum theory qua predictive framework. Given a suitable[1] expression for **H**, it is enough for all predictive purposes that (6.3.2) maintains a deterministic relationship between predictors (Hilbert space vectors) at different times. In any case, the **U**-evolution is that of the 'predictive potential' of a preparation, not that of a system and its physical attributes.

If a ket $|X>$ is expanded as (6.1.1), time-dependence is borne by the coefficients c_n, while the eigenkets $|u_n^A\rangle$ supply a fixed reference associated with definite values of the statistical distance $d_{n,X}$. No simple equation relates $d_{n,X}$ to its rate of change. On the other hand, from (6.1.1) and (6.3.2) :

$$i\, d_t c_k = \sum_n H^A_{kn} c_n , \qquad (6.3.3)$$

where H^A_{kn} is the matrix element $\langle u_k^A | \mathbf{H} | u_n^A \rangle$.

Any change in our knowledge brought forth by our gathering information through measurement has to be reflected in the formalism. Suppose the initial preparation is denoted by the ket $|X>$, expanded as (6.1.1) (with $c_i \equiv c_{i,A}$). Three different (highly idealised) cases will now be considered and contrasted:

Case1

An a_k-filter is set up after the emitting device or source of X-prepared samples (tests measurements of **A** are made to check that all outcomes a_i, $i \neq k$ are filtered out). Measuring **B** after the a_k-filter yields the outcome b_j, associated with the eigenvector $|v_j^B\rangle$, with probability

$$p^{(k)}(b_j) = p(b_j/a_k)p_X(a_k) = \left|\langle v_j^B | u_k^A \rangle\right|^2 \left|\langle u_k^A | X \rangle\right|^2 \qquad (6.3.4)$$

The action of the filter amounts to an effective a_k-preparation on which the probability of obtaining b_j is conditionalised. $p^{(k)}(b_j)$ is the product of the probability $p(b_j/a_k) = \left|\langle v_j^B | u_k^A \rangle\right|^2$ that measuring **B** on an a_k-filtered sample system will yield b_j and of the probability $p_X(a_k) = |c_k|^2$ that a released sample will pass the a_k-test. Weighting by the prior probability $|c_k|^2$ that an X-prepared sample will yield a_k if the sample is subjected to an **A**-measurement after the a_k-filter is no different from the weighting of the probability that a coloured ball drawn from vessel No. k will be blue, say, by the probability $|c_k|^2$ that the vessel (picked at random) is actually the k-th. The only presupposition is that the prior probabilities of **A**-measurement outcomes are not affected by the presence of an a_k-filter.

---

[1] See Section 2.2.

Case 2

Here, *no* a_k-filter is included in the set-up. The probability is now only conditional on the preparation at the source, i.e.

$$p_X(b_j) = \left|\langle v_j^B | X\rangle\right|^2. \qquad (6.3.5)$$

Regardless of the fact that no **A**-related filter is present, the predictive formalism gives us freedom to write that probability in terms of the eigenvectors of **A**:

$$p_X(b_j) = \left|\sum_i \langle u_j^B | u_i^A\rangle\langle u_i^A | X\rangle\right|^2. \qquad (6.3.6)$$

This is just an instance of the application of the α and β rules, since we can write, with the notation of Section 5.1:

$$p_X(X \to X_j^B) = |\sum_i A(X\to X_i^A)A(X_i^A\to X_j^B)|^2.$$

The β rule applies, as it should, in a situation where no a_i-filter is actually set up.

Case 3

The third and last case is one in which a *measurement* of **A** is performed on X-prepared samples, prior to **B** being measured. As in *Case 2*, no a_k-filter is part of the set-up. It is also assumed that a sample is not destroyed in the process of being **A**-measured. After a measurement of **A**, we shall suppose that a **B**-measurement is performed on the same sample, yielding b_j. The frequency of a given pair of outcomes (a_i,b_j) provides, up to fluctuations, an estimate of the probability $p(b_j \mid a_i)$ that the result a_i will be followed by b_j, both measurements being performed on the same sample. The probability of obtaining b_j whatever the outcome a_i is the sum over the possible a_i of the probabilities $p(b_j \mid a_i)$, each of them being weighted by the probability $p_X(a_i)$ that a prior measurement of **A** on the same (X-prepared) sample will yield a_i :

$$p_X(b_j) = \sum_i p(b_j \mid a_i)p_X(a_i) = \sum_i p_{i,A}(b_j) \mid c_i \mid^2 \qquad (6.3.7)$$

(6.3.7) is just the sum $\sum_i p^{(i)}(b_j)$, where each summand has the same form as (6.3.4), because $|c_i|^2$ appears in both expressions while $p(b_j \mid a_i)$ and $p(b_j/a_i)$ are evaluated using the same inner product. It is therefore tempting to *interpret* each $p^{(i)}(b_j)$ as an instance of the kind of situation from which (6.3.4) follows. Indeed, if one selects, in *Case 3*, only those pairs (a_k, b_j) for which measuring **A** gives the result a_k, the probability $p^{(k)}(b_j)$ is given

by (6.3.4). However, identifying with (6.3.4), *as obtained in Case 1*, the expression of $p^{(k)}(b_j)$ derived on the basis of such a selection supposes that picking the corresponding entries in the list of **A**-measurement outcomes is regarded as equivalent to the physical operation of an a_k-filter prior to measuring **B**, and that corresponding b_j entries in the **B**-list are understood as if they were results obtained in measuring **B** after such filtering. The identification is made despite the fact that no filter is included in the *Case 3* arrangement and the only preparation to which the prediction must refer to in that case is X.

Within the limits of the predictive scheme, **A**-measurements can be, although improperly, conceived as effecting **A**-filtering in a random fashion. There is no predictive harm in confusing with an actual eigenpreparation a merely conceptual sampling of outcomes as extracted from the **A**-list. Moreover, the mathematical identity of (6.3.4) and (6.3.7)-reduced-to-its-k-th-term comes in handy: if the result a_k of an **A**-measurement is known to be the case, calculating the probability of an outcome b_j of a subsequent **B**-measurement can be conveniently simplified using $p^{(k)}(b_j)$. The 'reduction' rule (**P7**) (see Section 4.2) amounts to taking advantage of a useful formal coincidence. In the words of Omnès[1],

> ...reduction is...a recipe allowing the erasure of irrelevant information: ...(one) forgets much of the initial preparation of the system (sometimes even everything). One sticks only to the result. The justification of this simplification is much less trivial in the present case than it is in mathematics and information theory, but the result is the same, namely, a *modus ponens* or essentially a rule for oblivion.

Unless A-measurements are made in the meantime, predictions from an X-preparation are computed using (6.3.5-6.3.6). In contrast, if measurements have actually been made, the correct expression for $p(b_j)$ is no longer (6.3.5) but (6.3.7). The computational difference lies in that, since (6.3.6) is the squared modulus of a sum, it contains non-quadratic 'cross-terms'. The change from (6.3.5-6.3.6) to (6.3.7) is usually interpreted as one from a pure case to a mixture. The word mixture conveys the idea that the simple sum of probabilities (6.3.7) just expresses one's ignorance of which outcome of the **A**-measurement is actually the case, it being understood there is a fact of the matter about any given outcome immediately after any **A**-measurement (one just has to look at the **A**-list to find out which amongst all possible outcomes is actually the case). However, since a pure case cannot consistently be regarded as describing a union of subensembles of physical systems endowed with given attributes[2], interpreting (6.3.5) as an expression of our ignorance is not a viable option. The essential difference between the two cases where (6.3.5) and (6.3.7) respectively hold is that, in the first, no

[1] Omnès 1994, p. 341.
[2] See section 4.3.

intermediate measurement is actually performed, nor are the conditions actually met for such a measurement. By contrast, in the second **A**-measurements are systematically performed prior to predicting the later occurrence of b_j. Therefore, the legitimacy of switching from (6.3.5) to (6.3.7) as the 'right' way of evaluating the probability $p(b_j)$ depends on the realisation of the appropriate context, associated here with the **A**-measurement. Switching from (6.3.5) to (6.3.7) raises none of the phenomenological issues that might be suggested by hasty parallels with a two-slit experiment. Whilst the presence of nonquadratic terms has a bearing on the value of probability, there does not correspond to it any observable effect such as a distinctive pattern of dark and bright fringes. Any 'interference' exhibited in (6.3.6) is, as it were, 'paper interference'. None the less, the change from (6.3.5) to (6.3.7) has often been described as a loss or disappearance of the cross-terms in (6.3.6). Implicit in that view is the belief that (6.3.5) and (6.3.7), and in particular the vectors entering in their evaluation, relate to objective physical attributes, measured or not, of an individual system. A sudden change in the form of probabilities would then reflect the occurrence of a real process, whereby any prior 'coherence' between the 'eigenstates' in superposition in (6.1.1) would be wiped out (as far, at least, as we can practically tell). However, as far as the predictive task is concerned – a task that is the sole purpose of the formalism outlined in this chapter and the next ones – it is idle to inquire about any 'losses'. Distinct operationally definite cases have to considered, to which there correspond distinct ways of calculating $p(b_j)$, and that is all. 'Interference' is just a characteristic of the Hilbert space-based formalism used to compute the probabilities. The β and the **PR** rule were shown to be derivable from consistency requirements and assumptions that nowhere make any reference to properties of physical systems or ensembles thereof. The β rule, in particular, holds in cases where no context is actually picked out that would supply valid statistical reference for the predictive task. On the other hand, one gets to (6.3.7) by assuming such a context *has* been picked out. The 'transition' from (6.3.5) to (6.3.7) can thus be understood as from a non-contextualised to a contextualised case, where each case calls for its own mode of probabilistic evaluation. Interference does not so much 'vanish' in the contextualised case as it simply does not arise: different circumstances, different uses of a single basic formalism. As to the 'reduction' of (6.3.7) to (6.3.4), it consists in nothing but updating the predictor as a result of acquiring the relevant 'outcome information'. The scheme does not (cannot) account for a given result being the case rather than another (among those with nonzero probability). But then, neither does the ordinary probability calculus (have to) account for a coin landing tails rather than heads in any single toss.

7

Hilbert space composition and quantum theory

7.1 n-ary preparations

Suppose that two different and totally unrelated preparations (I) and (II) are such that measuring **A** on any (I)-sample will yield a_m and no other a_k ($k \neq m$), and that measuring **B** on any (II)-sample will yield b_p and no other b_q ($q \neq p$). Letting $\left|u_i^A\right\rangle \in \mathbf{H}_A$ and $\left|v_k^B\right\rangle \in \mathbf{H}_B$ denote eigenkets of **A** and **B** respectively, the predictors for (I) and (II) can be simply 'pasted' by taking their direct product $\left|u_m^A\right\rangle\left|v_p^B\right\rangle = \left|\Psi_{m,p}^{A,B}\right\rangle \in \mathbf{H}_A \otimes \mathbf{H}_B$, where $\otimes$ is the symbol for the tensor product[1] of $\mathbf{H}_A$ and $\mathbf{H}_B$. The probability that measurements of **A** and **B**, performed on a (I)-sample and on a (II)-sample respectively, will yield a_k and b_l is given by

$$p_{\Psi_{m,p}}^{A,B}(a_k \wedge b_l) = \left|\left\langle u_k^A \right|\left\langle v_l^B \right|\left|\Psi_{m,p}^{A,B}\right\rangle\right|^2 \quad : \qquad (7.1.1)$$

the probability is calculated as if one were dealing with a single (I)+(II) preparation corresponding to the ket $\left|\Psi_{m,p}^{A,B}\right\rangle$.

There may not seem to be much benefit to be derived from such a representation, which is clearly equivalent to running the predictive algorithm using $\mathbf{H}_A$ and $\mathbf{H}_B$ kets separately. Furthermore, since a complete set of $\left|\Psi_{m,p}^{A,B}\right\rangle$ form a basis of eigenkets of $\mathbf{A}\otimes\mathbf{B}$, any ket $\left|\Psi\right\rangle \in \mathbf{H}_A \otimes \mathbf{H}_B$ can be expanded as

$$\left|\Psi\right\rangle = \sum_{i,k} c_{ik}\left|\Psi_{i,k}^{A,B}\right\rangle = \sum_{i,k} c_{ik}\left|u_i^A\right\rangle\left|v_k^B\right\rangle \qquad (7.1.2)$$

Can any sense be made of, and any use be found to a notion of *binary* preparation, or more generally of *n-ary* preparation, to which a ket in the tensor product $\mathbf{H}_A \otimes \ldots \otimes \mathbf{H}_K$ of *n* Hilbert spaces would be associated? Besides the two trivial instances of such preparations we have already encountered: (i) *unary* (n=1) and (ii) that which boils down to a mere conjunction ('pasting') of n preparations (product of *n* kets), it is by no means obvious whether any other can be reasonably conceived, to which a non-trivial linear combination (7.1.2) of ket products would be ascribed. On the other hand, if what matters to prediction is the statistical distance between the current preparation and suitably chosen reference eigenpreparations,

[1] If dim ($\mathbf{H}_A$) = n_A and dim($\mathbf{H}_B$) = n_B , then dim($\mathbf{H}_A \otimes \mathbf{H}_B$) = $n_A \times n_B$.

a complete set of eigenket products $|\Psi^{A,B}_{i,k}\rangle = |u^A_i\rangle|v^B_k\rangle$ is a natural choice for providing such a reference. If $p_\Psi(a_m,b_p)$ is the probability of obtaining results a_m and b_p when a joint measurement of **A** and **B** is carried out on 'Ψ-prepared' samples, where $|\Psi\rangle$ has the general form (7.1.2), then

$$p_\Psi(a_m,b_p) = \left|\langle\Psi^{A,B}_{m,p}|\Psi\rangle\right|^2 = \cos^2 d(\Psi,\Psi^{A,B}_{m,p}) = |c_{ik}|^2 \qquad (7.1.3)$$

With the notation of Section 6.2, $d(\Psi,\Psi^{A,B}_{m,p})$ is the statistical distance between the Ψ preparation associated with $|\Psi\rangle \in \mathbf{H}_A \otimes \mathbf{H}_B$ and the reference eigenpreparation $\Psi_{m,p}{}^{A\text{-}B}$ (the latter is in fact nothing but the conjunction of two independent preparations). Let us temporarily set aside the question of the meaning of such a binary preparation and focus on the possible significance of such a predictive vector.
If (7.1.2) reduces to a simple product $|\Psi> = |u>|v> \in \mathbf{H}_A \otimes \mathbf{H}_B$ with $|u> \in \mathbf{H}_A$ and $|v> \in \mathbf{H}_B$, then

$$p_\Psi(a_m,b_p) = p_u(a_m)p_v(b_p), \qquad (7.1.4)$$

where $p_u(a_m) = \left|\langle u^A_m|u\rangle\right|^2$ and $p_v(b_p) = \left|\langle v^B_p|v\rangle\right|^2$: this probability is the product of the probabilities for the outcomes as computed separately for each preparation. Since (7.1.4) is a 'quantal' version of the classical rule for the computation of the joint probability of two independent events, it may be worthwhile to consider whether the classical formula

$$p(a_m \wedge b_p) = p(a_m).\ p(b_p/a_m)\ , \qquad (7.1.5)$$

where $p(b_p/a_m)$ is the probability of b_p conditional on a_m being the case, has any quantum-theoretical analogue. When the 'events' are independent, (7.1.5) reduces to a product of probabilities, in direct analogy to (7.1.4). In the ordinary probability calculus, the probability of the outcome b_p obtains via summing over the complete set of the a_i:

$$p(b_p) = \sum_i p(a_i).p(b_p/a_i) \qquad (7.1.6)$$

The closest we can get, in the linear scheme, to the classical formulae (7.1.5-7.1.6) follows from a simple algebraic property: any ket $|\Psi>$ of the form (7.1.2) can be written

$$|\Psi\rangle = \sum_i \lambda_i |u^A_i\rangle|w^B_i\rangle \qquad (7.1.7)$$

where the $|w^B_i\rangle$ are normalised $\mathbf{H}_B$ vectors. This so-called *Schmidt* or *biorthogonal decomposition* of the ket $|\Psi>$ secures a one-to-one relationship

between $\mathbf{H}_A$ and $\mathbf{H}_B$ terms by the association of a *relative*[1] ket $|w_i^B\rangle$ with any given ket $|u_i^A\rangle$. The mutually orthogonal $|w_i^B\rangle$ may then be regarded as eigenkets of some observable $\mathbf{R}_{\Psi/A}$ operating on $\mathbf{H}_B$. For example, the ket

$$|\Psi> = \frac{1}{2}|z_+>|z_+> + \frac{1}{2}|z_+>|z_-> + \frac{1}{2}|z_->|z_+> - \frac{1}{2}|z_->|z_-> \qquad (7.1.8)$$

where $|z_\pm>$ are eigenkets of the spin observable σ_z corresponding to the eigenvalues ±1, can be written in biorthogonal form as

$$|\Psi> = 2^{-1/2}|z_+>|x_+> + 2^{-1/2}|z_->|x_->, \qquad (7.1.9)$$

where $|x_\pm>$ are eigenkets of σ_x ($\equiv \mathbf{R}_{\Psi/\sigma z}$) associated with the eigenvalues ±1 ($\dim(\mathbf{H}_{\sigma_x}) = \dim(\mathbf{H}_{\sigma_z}) = 2$). The decomposition (7.1.7) is not unique – (7.1.9) is a case in point – unless the moduli of the coefficients are all distinct.

If each of the products $|\Phi_i\rangle = |u_i^A\rangle|w_i^B\rangle$, considered separately, is interpreted as relating to a conjunction of two preparations (I) and (II), these will be such that the outcomes of **A**-measurements on (I) and $\mathbf{R}_{\Psi/A}$-measurements on (II) are uncorrelated. The probability that a_m and b_p will obtain when **A** and **B** are independently measured is

$$p_\Psi(a_m \wedge b_p) = |\langle u_m^A|u_i^A\rangle|^2 |\langle v_p^B|w_i^B\rangle|^2 \qquad (7.1.10)$$

On account of the $|u_i^A\rangle$ being normalised and mutually orthogonal, (7.1.10) is non-zero only if i = m. This suggests interpreting $|\langle v_p^B|w_i^B\rangle|^2$ as the probability of obtaining b_p, conditional on the preparation being such that measuring **A** will yield a_i with probability 1. Then, $p_{\Phi_i}(a_i \wedge b_p) = p_{\Phi_i}(a_i).p_{\Phi_i}(b_p/a_i)$, with $p_{\Phi_i}(a_i) = 1$. Using (7.1.7) and summing over the *i*,

$$p_\Psi(b_p) = \sum_i |\lambda_i|^2 |\langle v_p^B|w_i^B\rangle|^2 = \sum_i p_\Psi(a_i).p_{\Phi_i}(b_p/a_i), \qquad (7.1.11)$$

where $p_\Psi(a_i) = |\lambda_i|^2 = \sum_k |c_{ik}|^2$.

(7.1.11) looks similar enough to the classical total probability rule (7.1.6). However, the expression essentially differs from the latter by the nonexistence of such homogeneous 'background conditions' as classical treatment assumes: $p_{\Phi_i}(b_p/a_i)$ explicitly refers to a preparation Φ_i for which measuring **A** would yield a_i with probability equal to 1, whereas $p_\Psi(a_i) \neq 1$ when $|\Psi>$ does

[1] The idea was introduced under the name 'relative state' by Everett 1957. Needless to say, switching from (7.1.2) to (7.1.7) implies no acceptance of Everett's interpretation, let alone of any of its many-worlds/minds offshoots.

not reduce to any one **A**-eigenket. Each summand on the RHS of (7.1.11) expresses, as it were, a conditionalisation of b_p on the a_i with respect to a 'virtual' i.e. counterfactual eigenpreparation Φ_i. Contributions of all Φ_i must be taken into account in the calculation of $p_\Psi(b_p)$, just as a complete set of non-actualised 'intermediate links' must be in the calculation of the total amplitude using the β rule. Another fundamental limitation of the analogy is that the uniqueness of the Schmidt decomposition is not guaranteed for all[1] kets $|\Psi>$. In what follows, we shall just acknowledge the usefulness of a tensor product extension of the basic formalism outlined in chapter 6, postponing the question of how to justify tensor product composition till chapter 9. That kind of extension is exploited in quantum theory in two basic circumstances:

(i) when two *compatible* observables are to be jointly measured on samples from a *unary* preparation.

An instance of (i) would be the assessment of both spatial and spin degrees of freedom on identically prepared samples. Consider a Stern-Gerlach experiment in which a beam of silver atoms is subjected to a non-uniform magnetic field in the z direction. Since the experiment gives rise to two distinctive regions of impact on a screen in the z direction, observed values of the spin component[2] σ_z are then correlated to those of the spatial z coordinate.

(ii) When two observables, compatible or not, are measured on the members of *pairs* (n-uples) generated in the process of a genuinely *binary* (n-ary) preparation.

Instances of (ii) are the generation of two-photon 'cascades[3]' or the splitting of a diatomic molecule. We shall admit that there exist experimental situations that give rise to a 'fragmented preparation', on the members of which it is then possible to act separately. However, given our purely predictive aim, we shall not be concerned with the thorny matter of deciding whether such 'fragmentation' can legitimately be analysed into genuine 'particles'.

So far, Hilbert spaces $\mathbf{H}_A$, $\mathbf{H}_B$ etc. have been labelled with reference to particular observables. This does not mean, of course, that predictions are

---

[1] In the ternary case, it has been proved (Bub 1999, pp. 164-167) that the triorthogonal decomposition of a ket $|\Psi> = \sum_i c_i |u_i> |v_i> |w_i>$, when it exists, is unique without any condition on the coefficients. Such a decomposition, however, does not always exist. The issue remains unsettled for n>3.

[2] For any direction α, the σ_α notation for spin-½ operators is used rather than $\mathbf{S}_\alpha = ½\,\sigma_\alpha$, and the ubiquitous constant ∞ is set equal to 1.

[3] As in the Paris XI-Orsay experiments of Aspect *et al.*1982.

restricted to a set of 'preferred' observables. Rather, the choice of **A** and **B** sets the dimension, finite or not, of the relevant Hilbert space. On the other hand, since we shall be mainly concerned with (ii), a distinction must be clearly maintained between experiments performed separately on either member (1) or (2) of any produced 'pair'. Accordingly, the relevant Hilbert spaces will from now on be designated as $\mathbf{H}_1$ and $\mathbf{H}_2$. One should, in any case, refrain from the temptation to conceive of either Hilbert space as a 'state space' for subsystems that, together, would constitute the sample pairs prepared as 'specified' by a given $\mathbf{H}_1\otimes\mathbf{H}_2$ ket. Any reference to composite systems and subsystems allegedly involved in *n*-ary preparations will, in the remainder of this chapter, be fundamentally conventional or 'symbolic'. The discussion will also be almost exclusively confined to binary preparations, bearing in mind that all remarks and conclusions generally apply to *n*-ary (n>1) preparations. Any Hilbert spaces involved will be thought of as referring to potential choices of observables to be measured on sample pairs or pair members separately. For example, $\mathbf{H}_2$ may be chosen as the two-dimensional linear space appropriate to predicting outcomes of measuring any spin (½) component σ_n; the subscript 2 here indicates only that the corresponding information is restricted to 'member 2' of any sample pair. It is only required that this label be consistently maintained for a given set-up and a given sequence of measurements (e.g. those detections or measurements that occur 'on the left', as opposed to those occurring 'on the right', may conventionally be labelled '1').

7.2 Correlations and entanglement

The probability of obtaining result a_m if an **A**-measurement is made on member 1 of a sample pair, given that all such pairs originate in a single binary preparation associated with the statistical operator ρ, is given by

$$p_\rho(a_m^{(1)}) = \mathrm{tr}(\rho(\mathbf{P}_{A,m}\otimes\mathbf{I}) \qquad (7.2.1)$$

where the trace is taken over any complete orthonormal set in $\mathbf{H}_1\otimes\mathbf{H}_2$, $\mathbf{P}_{A,m}$ projects onto the ray associated with the eigenvalue a_m of **A** and **I** is the identity operator[1]. The probability of joint results a_m and b_p of an **A**-measurement on 1 *and* a **B**-measurement on 2 is

$$p_\rho(a_m^{(1)},b_p^{(2)}) = \mathrm{tr}(\rho(\mathbf{P}_{A,m}\otimes\mathbf{P}_{B,p})). \qquad (7.2.2)$$

If gathering information on 1-samples has no bearing whatsoever on 2-samples (and conversely), then ρ can be written

[1] In any tensor product $\mathbf{X}\otimes\mathbf{Y}$, **X** (left) is taken to act upon $\mathbf{H}_1$, and **Y** (right) upon $\mathbf{H}_2$.

$$\rho = \rho^{(1)} \otimes \rho^{(2)} \qquad (7.2.3)$$

where ρ , $\rho^{(1)}$ and $\rho^{(2)}$ all correspond to pure cases. (7.2.3) is then equivalent to the ket product $|\Psi> = |u_1>|u_2> \in \mathbf{H}_1 \otimes \mathbf{H}_2$, where $|u_i> \in \mathbf{H}_i$ (all density operators involved are projectors: $\rho^{(i)} = |u_i><u_i|$ and $\rho = |\Psi><\Psi|$). Each $\rho^{(i)}$ (i=1,2) obtains (trivially) by performing a partial trace over $\mathbf{H}_j$ (j = 3-i): $\rho^{(i)} = tr_j(\rho)$. Partial tracing can be thought of as averaging over any complete set of eigenkets of $\mathbf{H}_j$, hence over all eigenpreparations that may be used as a reference for assessing statistical distance with respect to 'subsystem' j. The partial trace uniquely determines, for any ρ (whether or not it takes the product form (7.2.3)), two *reduced* or *marginal* statistical operators $\rho^{(i)}$ that can be used to calculate the probabilities of the outcomes of measurements to be performed on i-samples (i = 1 or 2) separately. These two operators must satisfy an evident consistency constraint: as long as predictions are restricted to outcomes of measurements performed on i-samples exclusively, they must agree whether they are calculated using ρ or $\rho^{(i)}$, e.g. with i=1:

$$p_\rho(a_m^{(1)}) = tr(\rho \mathbf{P}_{A,m} \otimes \mathbf{I}) = tr_1(\rho^{(1)} \mathbf{P}_{A,m}) = p_{\rho^{(1)}}(a_m). \qquad (7.2.4)$$

Equivalently, $\rho^{(i)}$ must give rise to the same expectation value as ρ for all observables measured on i-samples only.

Most[1] binary preparations will be associated with non-trivial linear combinations (7.1.2). As a result, pure case statistical operators will *not* in general reduce, via partial tracing, to $\rho^{(i)}$ which are pure. Most importantly, if the reduced statistical operators are 'mixed' i.e. expressible as irreducible linear combinations of projection operators, then these operators do *not* suffice to determine the statistical operator ρ of the binary preparation. This is usually interpreted as an indication that the preparation embodies 'more information', in the form of correlations empirically accessible via joint measurements of sample pairs, than can reveal measurements performed on 1-samples or 2-samples only. The *correlation index* a.k.a. 'mutual information[2]'

---

[1] The set of product kets is a zero-measure subset of $\mathbf{H}_1 \otimes \mathbf{H}_2$.

[2] The term originates in classical (Shannon) information theory, where the joint entropy H[S,S'] is at most equal to the number H[S] + H[S'] of binary questions that are needed to find out the content of two messages S and S' (H is the Shannon information function or 'entropy'). It is strictly lower if the two messages are not statistically independent. If the S-message has been figured out, the number of questions it will take to find out S' is the entropy of S', conditional on S: H[S'/S] = H[S,S'] – H[S]. The 'mutual information' I[S,S'] = H[S'] – H[S'/S] is the amount we learn about S' from knowing S, i.e. it corresponds to the average number of questions that knowing S saves us from having to ask about S'. Thus if S' represents the signal picked up by the receiver, I[S,S'] quantifies how much of the emitted message S is, as it were, embodied in that signal. The qualification 'mutual' arises just because I happens to be symmetrical in its arguments: I[S,S'] = I[S',S]. However, mutual

$$I_c = S(\rho^{(1)}) + S(\rho^{(2)}) - S(\rho) \qquad (7.2.5)$$

where $S(\rho) = -tr(\rho log \rho)$ is the *von Neumann 'entropy'*[1], quantifies the possibility of any information gain over and above that provided by measurements performed on 1 and 2 separately (without any comparison of lists of 1 and 2-data). $I_c = 0$ only if ρ can be written in the product form (7.2.3) i.e. in the absence of any correlation. In that case, the expectation value for joint outcomes of measuring **A** on 1-samples and **B** on 2-samples is a simple product of the two means obtained separately:

$$< \mathbf{A}\otimes\mathbf{B} >_\rho = tr(\rho(\mathbf{A}\otimes\mathbf{B}) = tr(\rho^{(1)}\mathbf{A}).tr(\rho^{(2)}\mathbf{B}) = < \mathbf{A} >_{\rho^{(1)}} < \mathbf{B} >_{\rho^{(2)}} . \qquad (7.2.6)$$

Conversely, if ρ cannot be written in product form, then there must be correlations ($I_c \neq 0$). For example, if a mixed case binary preparation is such that $\rho_S = \sum_k w_k(\rho_k^{(1)} \otimes \rho_k^{(2)})$, with $w_k \in [0,1]$ and $\sum_k w_k = 1$ (with more than one $w_k \neq 0$), then the index $I_c \neq 0$, or

$$< \mathbf{A}\otimes\mathbf{B} > = tr(\rho_S(\mathbf{A}\otimes\mathbf{B}) = \sum_k w_k\, tr(\rho_k^{(1)}\mathbf{A})\, tr(\rho_k^{(2)}\mathbf{B}) \quad , \qquad (7.2.7)$$

which is clearly different from (7.2.6). Since such a 'separable' mixed case is just a binary version of the sort of ignorance-interpretable mixture exemplified by (4.4.1), any correlations can be given a common cause explanation.

Let us now consider the biorthogonally expanded ket

$$|\Psi\rangle = \sum_i c_i |u_i^A\rangle |v_i^B\rangle , \qquad (7.2.8)$$

---

information is a rather unfortunate expression that should rather be avoided, for it suggests that subsystems, qua potential sources, may hold information about one another, although there is no clear notion of what 'holding information' could mean for physical systems if no reference is made to non-physical notions such as pattern recognition abilities etc.

[1] The (**PR**) expression of probability $p_k = tr(\rho \mathbf{P}_k)$ being inserted in Shannon's function $H_S = -\Sigma_k p_k log p_k$ (Shannon 1948 ; Cover and Thomas 1991), optimal predictability obtains (Wehrl 1978) when the projection operators $\mathbf{P}_k$ form a set of eigenprojectors for the statistical operator ρ. The 'quantum version' of H_S then reduces to $S(\rho)$, which quantifies the predictive uncertainty associated with a standard projective measurement in optimal conditions. Predictability is maximal in the pure case, for which $S(\rho) = 0$, and least in the 'maximally mixed' case (ρ proportional to the identity operator) for which the outcomes of any measurement are equally likely. $S(\rho)$ can be thought of as measuring the minimal amount of information that is missing towards identification of a preparation as a pure case, given any mixed case statistical operator ρ. Its value cannot exceed the logarithm of the dimension of the Hilbert space i.e. of the least number of orthogonal vectors that can form a basis for that space. It should be noted that whether $S(\rho)$ qualifies as a quantum-theoretical analogue of the thermodynamical entropy remains controversial (Shenker 1999, Hemmo and Shenker 2003).

where $|u_i^A\rangle$ and $|v_i^B\rangle$ are eigenkets of two observables **A** and **B** and more than one c_i is non-zero. A predictive vector like (7.2.8), or more generally (7.1.2), does not generally factor into $\mathbf{H}_1$ and $\mathbf{H}_2$ kets. Erwin Schrödinger[1] coined the adjective *entangled* to designate such non-factorability. In his first characterisation of so-called *entanglement*, his emphasis is, significantly, not on any changes in some definite 'states' of the bodies involved, but on structural changes in our 'maximal knowledge' as afforded by the relevant Ψ-function or Hilbert space vector (7.1.2):

> If two separated bodies, each by itself *known maximally*, enter a situation in which they influence each other, and separate again, then...the *knowledge remains maximal*, but at its end, if the two bodies have again separated, it is *not* again split into a *logical sum of knowledges* about the individual bodies[2].

In order to shed light on the peculiar traits and possible difficulties alluded to by Schrödinger, let us consider the density operator $\rho = |\Psi><\Psi|$, where $|\Psi>$ is given by (7.2.8). Then,

$$\rho = \sum_i |c_i|^2 |u_i\rangle\langle u_i| \otimes |v_i\rangle\langle v_i| + \sum_{i\neq k} c_i c_k^* |u_i\rangle\langle u_k| \otimes |v_i\rangle\langle v_k| . \qquad (7.2.9)$$

The first sum in (7.2.9): $\rho_d = \sum_i |c_i|^2 |u_i\rangle\langle u_i| \otimes |v_i\rangle\langle v_i| = \sum_i |c_i|^2 \rho_i^{(1)} \otimes \rho_i^{(2)}$ is just of the same form as the mixed case statistical operator ρ_S we considered above. Since the second, non-diagonal or 'interference' term in (7.2.9) gives a null contribution to the partial traces, the reduced statistical operators

$$\rho^{(1)} = tr_2\rho = \sum_i |c_i|^2 |u_i\rangle\langle u_i| \quad \text{and} \quad \rho^{(2)} = tr_1\rho = \sum_i |c_i|^2 |v_i\rangle\langle v_i| \qquad (7.2.10)$$

do not differ in any way from those derived from ρ_d alone. This means that all the statistics to be derived from the $\rho^{(i)}$ must fail to distinguish[3] between ρ and ρ_S. No series of measurements performed on 1-samples alone, or 2-samples alone, can tell ρ_S and ρ apart; nor can the statistics of any measurements performed on 1 and 2-samples, the 'joint' observables being of the form $\mathbf{U}\otimes\mathbf{V}$, where **U** commutes with **A** and **V** commutes with **B**. On the other hand, ρ and ρ_S can in principle be distinguished by performing suitable measurements on both 1 and 2 (the relevant observables should be linear combinations of tensor products of $\mathbf{H}_1$ and $\mathbf{H}_2$ observables, or such

---

[1] Schrödinger 1935a.

[2] Schrödinger 1935b (italics added). However, Schrödinger occasionally reverts to more problematic state-based formulations, as in his discussing the EPR issue: "It is rather discomforting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter's mercy in spite of his having no access to it." (Schrödinger 1936, p. 451)

[3] Von Neumann 1955, pp. 426-429.

products $\mathbf{X}\otimes\mathbf{Y}$ provided $\mathbf{X}$ does not commute with $\mathbf{A}$ and $\mathbf{Y}$ does not commute with $\mathbf{B}$).

If the reduced density operators $\rho^{(i)}$ underdetermine the potential for prevision that is embodied in a binary preparation described by ρ (unless one of them at least is a projection operator), correlations amongst partitions of a n-ary preparation suffice to determine the associated statistical operator. Given a n-ary preparation to which there corresponds a density operator ρ, then ρ is entirely determined by the mean values $< \mathbf{X}\otimes\mathbf{Y} >_\rho = \mathrm{tr}(\rho(\mathbf{X}_1\otimes\mathbf{X}_2\ldots\otimes\mathbf{X}_n))$, where $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$ are suitably chosen observables acting upon $\mathbf{H}_1$, $\mathbf{H}_2,\ldots\mathbf{H}_n$ respectively: everything that may be predicted given a n-ary preparation is completely pinned down by the expectation values of $\mathbf{H}_1\otimes\mathbf{H}_2\otimes\ldots\otimes\mathbf{H}_n$ observables, all of which are products of $\mathbf{H}_1$, $\mathbf{H}_2,\ldots\mathbf{H}_n$ observables. In other words, the relevant statistical operator can always be inferred from the correlations among results of measurements performed on each subsystem separately[1]. This basic property of the formalism follows from the observation that a basis for the algebra of $\mathbf{H}_1\otimes\mathbf{H}_2$ observables is provided by a subset of the set of all products of $\mathbf{H}_1$ and $\mathbf{H}_2$ observables, together with the fact that the algorithm whereby expectations values are computed is linear on the algebra of observables[2]. All the correlations for any m-partite resolution of a n-ary preparation (e.g. into $\mathbf{H}_1$ and $\mathbf{H}_2\otimes\ldots\mathbf{H}_n$, into $\mathbf{H}_1\otimes\mathbf{H}_2$, $\mathbf{H}_3\otimes\ldots\otimes\mathbf{H}_k$ and $\mathbf{H}_{k+1}\otimes\ldots\otimes\mathbf{H}_n$, etc.) can be calculated from the appropriate statistical operator ρ on $\mathbf{H}_1\otimes\mathbf{H}_2\otimes\ldots\mathbf{H}_n$. Moreover, these are all the correlations ρ can possibly encapsulate (Gleason's theorem[3]).

Might tensor product composition itself follow from requiring previsional consistency amongst all possible resolutions? Some recent work[4] hints at this possibility. Meanwhile, similar considerations have motivated the suggestion[5] ('Ithaca Interpretation') that quantum theory might after all be fundamentally about correlations rather than correlata ('states' and/or attributes of individual systems). Since "anything you can say in terms of quantum states – and some strange things can be stated in that language – can be translated into a statement about subsystem correlations i.e. about joint probability distributions[6]", a ket or density operator in the n-ary case appears to be "nothing more than a concise encapsulation of the correlations

---

[1] Wootters 1990; Mermin 1998, Appendix 1. For example, the $15 = 4^2 - 1$ means $< \sigma_i\otimes\mathbf{I} >$, $< \mathbf{I}\otimes\sigma_j >$ and $< \sigma_i\otimes\sigma_j >$ (i,j=x,y,z) entirely determine the density operator for any binary ('two-qubit') preparation.

[2] For instance, the density operator for the singlet is expanded in the operator basis $\{\mathbf{I}_1\otimes\mathbf{I}_2, \sigma_x\otimes\sigma_x, \sigma_y\otimes\sigma_y, \sigma_z\otimes\sigma_z\}$ as $\rho = \frac{1}{4}(\mathbf{I}\otimes\mathbf{I} - \sigma_x\otimes\sigma_x - \sigma_y\otimes\sigma_y - \sigma_z\otimes\sigma_z)$.

[3] Gleason 1957.

[4] Fuchs 2002 ; see Section 9.3.

[5] Mermin 1998.

[6] *Ibid.*

among its subsystems[1]." The kind of linear predictive scheme outlined in this and the previous chapter encompasses all the relationships between the statistics of measurements for all predictively relevant resolutions of a n-ary preparation. Correlations can be revealed by a careful comparison of outcome lists for measurements made independently on matching pair or n-uple members. Quantum teleportation or dense coding, discussed in chapter 8, provide a most striking illustration of how the potential to exhibit correlations in statistics of measurement outcomes that is embodied in some n-ary preparations may be brought into use, but also of the risks inherent in forcing objectivistic and non-statistical readings upon the interpretation of the data.

Operators that represent symmetry transformations of Hilbert space must be (anti)unitary[2]. Since the tensor product of two (anti)unitary operators is an (anti)unitary operator acting upon vectors of the composite Hilbert space[3], (anti)unitary transformations operating separately in $\mathbf{H}_1$ and $\mathbf{H}_2$ have a definite net effect on elements of $\mathbf{H}_1 \otimes \mathbf{H}_2$. On the contrary, the $\otimes$ product of a unitary and an antiunitary operation results in a transformation of $\mathbf{H}_1 \otimes \mathbf{H}_2$ that is neither unitary nor antiunitary with respect to a binary preparation; it is indefinite unless the preparation can be represented as a product of kets or a separable mixture ρ_S : the net effect of the combined transformation on the latter is another statistical operator that is also separable. Now, any antiunitary transformation can be written as the product of a unitary operator and the *time reversal* operator. Accordingly, the time evolution of a two-qubit separable statistical operator (or factorable ket) can be factored into a product involving 'time arrows' pointing towards opposite directions. The temporal structures of the $\mathbf{H}_1$ and $\mathbf{H}_2$ contributions to a separable mixture are thus disconnected: it is as if the whole does not keep track of a single time direction in respect of both arms of the preparation, so that time reversal 'locally' affects that part of the predictor that is relevant to one arm, without affecting the time evolution of the other[4]. On the other hand, an *in*separable preparation is such that an

---

[1] *Ibid.*

[2] Wigner 1959, p.233.

[3] A tensor product of operators (matrices) is unitary if and only if each one of the operators (matrices) is unitary up to a constant (see e.g. Hungerford 1974).

[4] Sanpera *et al.* 1997. Reversing time on one of the arms of a two-qubit preparation is equivalent to complex conjugation being taken on the relevant part of Hilbert space. This operation yields the partial transpose of the density matrix ρ. For the directionality of time in subsystems to remain uncorrelated, the transpose must satisfy the Peres separability criterion (Peres 1996) i.e. the matrix obtained after partial transposition of the total density matrix ρ must have non-negative eigenvalues only ($\rho^{T}{}_2 = (\rho^{T}{}_1)^* \geq 0$). The condition is both sufficient and necessary provided that the dimension of the tensor product Hilbert space does not exceed 2×2 (two-qubit case) or 2×3 (qubit-trit case). For composite systems associated with spaces of higher dimensions, the positivity condition of $\rho^{T}{}_2$ is only a necessary one for separability: statistical operators have been found that satisfy the criterion despite their not being separable.

antiunitary transformation performed on any of its subsystems would jeopardise the time evolution of the whole. This suggests regarding entanglement as required to maintain consistency in temporal evolution of the full predictive potential of a n-ary preparation. On the other hand, one should refrain from thinking of entanglement as indispensable for the manifestation of 'non-classical' statistics. Sets of orthogonal ket products relating to two 'three-level' particles have been exhibited[1], which cannot be reliably distinguished by two arbitrarily separated observers who are given one particle each, as long as the observers do not know which preparation, associated with a ket in the set, they actually deal with (as usual in QIT Gedankenexperiments, the observers are allowed to perform only local operations – i.e. on their own particles in their own labs – and to communicate using classical means, e.g. a phone line). The existence of such possibilities should warn us against a tendency to over-emphasise the importance of entanglement.

7.3 Binary case objectification

Given a large ensemble of N pairs to the preparation of which a ket (7.1.2-7.2.8) is ascribed, suppose it is claimed that a definite value a_m of **A** pertains, prior to any measurement, to a number

$$N_m^{(1)} \approx N p_{\rho^{(1)}}(a_m) = N\mathrm{tr}(\rho^{(1)}\mathbf{P}_{A,m}) = N|c_m|^2$$

of the 1-subsystems in the ensemble, and that a definite value b_p of **B** similarly pertains to a number $N_p^{(2)}$ of the 2-subsystems. An ensemble $\mathbf{E}_{m,p}$ of 1+2 pairs whose members have those attributes should then, following d'Espagnat[2], be correctly described using the ket product $|u_m\rangle\otimes|v_p\rangle \in \mathbf{H}_1\otimes\mathbf{H}_2$, or equivalently the density operator $\rho_{m,p} = |u_m\rangle\langle u_m| \otimes |v_p\rangle\langle v_p|$. If all there is to the preparation is an ordinary mixture of all such pairs, then the whole ensemble of pairs should be the set-theoretical union $\bigcup_{m,p} \mathbf{E}_{m,p}$ of all such subensembles. But a statistical operator that corresponds to that union is a weighted sum of all the $\rho_{m,p}$, hence a separable density operator that can in principle be distinguished from (7.2.9). Therefore, mixed cases represented by density operators $\rho^{(1)}$ and $\rho^{(2)}$ (obtained by partial-tracing ρ) cannot be regarded as descriptions of ordinary mixtures that would reflect one's ignorance of actually possessed values. 'Mixtures' (7.2.10) obtained by partial tracing from (7.2.8) are, appropriately enough, sometimes called improper[3], as

[1] Horodecki *et al.* 1999, Bennett *et al.* 1999.
[2] d'Espagnat, 1976, pp.59-61; 1995, 7.3
[3] d'Espagnat 1976.

opposed to 'proper' mixtures of which an interpretation can be given in terms of attributes that are genuinely possessed but (yet) unknown.

The trouble with objectification in the quantum setting is that one considers as reflections of attributes of individual objects ('particles') values that can be ascertained only in a piecewise manner in definite contexts, the realisation of which calls for mutually exclusive experimental arrangements. If empirical access to measured values of physical quantities is so constrained, it is little wonder any essentially context-independent prior value assignment should conflict with the predictions of quantum theory[1]. That no 'instruction set' or simple common cause explanation, associated with an 'individual particle' view of pair members, can possibly account for such correlations is just a reminder that a predictive scheme in which statistical distance appears to play a central role is irreducible to a simple-minded ensemble-based picture, with all the 'context-free' expectations that usually accompany such a picture.

Entanglement has recently come to the fore with the rise of Quantum Information Technology (QIT). Quantum teleportation, (super)dense coding, quantum algorithms and quantum cryptography take advantage of hitherto unsuspected data storage and processing capacities implied by Hilbert space structure and tensor product composition[2]. Entanglement also helps to protect information by spreading it over a 'N-qubit' codeword or block. DiVincenzo and Peres[3] have used 5-qubit[4] blocks to give one more illustration of the conflict of the SAQM with some unwarranted expectations. The claim they consider is that indirect ways of determining the value of, say, σ_x for one of the qubits by measuring other qubits in the block must necessarily agree – as if that value were decided *a priori*, regardless of any choice and actual implementation of a measuring procedure whereby such a value is ascertained. They show that such a claim cannot be reconciled with the SAQM. The proof is nothing but a 32-dimensional equivalent to earlier ones involving 2 or 3-particle entanglement[5].

On the purely predictive view of quantum theory that is developed here (Part B) there is no 'measurement problem', since the latter is an offshoot

---

[1] Following Greenberger *et al.* 1990, Mermin 1990 gives a simple and elegant algebraic proof that any such value-assignment leads has to conflict with the SAQM. The proof involves three spin-½ fermions and crucially hinges on the anticommutativity ('incompatibility') of the Pauli operators.

[2] See Clifton 2001 for a review of basic issues raised by entanglement in QIT; Lo *et al.* 1998.

[3] DiVincenzo and Peres 1996.

[4] Quantum error-correcting codes take advantage of the resilience of N-qubit blocks under random corruption of part of their encoding. Being able to correct general errors requires at least 5 qubits to encode any one qubit (Knill and Laflamme 1997); for a review, see J. Preskill, in Lo *et al.* 1998, p.224., and Section 8.1 for the definition of a qubit.

[5] Mermin 1990.

of the basic difficulties raised by objectification in the binary case[1]. Within the extended framework outlined in this chapter, values a_k of a system observable **A** may be correlated with definite values m_k of an ad hoc 'pointer observable' through some appropriate correspondence between $\mathbf{H}_A$ and $\mathbf{H}_M$ kets (expanded in the $\{|a_i>\}$ and the $\{|m_i>\}$ basis respectively). Starting from an initial ket $|\varphi_0> \in \mathbf{H}_A$ and 'pasting' it to a 'void' ket $|m_0>$: $|\varphi_0>|m_0> \in \mathbf{H}_A \otimes \mathbf{H}_M$, a suitable unitary evolution of that product can be worked out, such that a strict $a_k \leftrightarrow m_k$ correlation allows pointer readings m_k to be regarded as faithful representatives of measured **A** values a_k. The **U**-evolved ket $|\Psi_{S+M}> = \sum_j c_j |a_j\rangle|m_j\rangle$ can be treated as if it were associated to a genuine binary preparation involving system and apparatus. The one-to-one $a_k \leftrightarrow m_k$ correspondence embodied in the predictor $|\Psi_{S+M}>$ guarantees that predictions for system-related observables are reflected in the apparatus i.e. in the $\mathbf{H}_M$ contribution. There is little if any gain in introducing the extra set $\{|m_i>\}$, for this merely 'doubles' the mathematical resources needed for the calculation of outcome probabilities.

[1] Mittelstaedt 1997.

8

Quantum Teleportation
a case study

8.1. 'Teleporting a quantum state'

The seemingly puzzling claim has recently been made[1] that, even if nothing is known about the quantum 'state' of a given individual system, a perfect replica of that state can none the less be made available 'at a distance'. The state would be "disassembled[2]" at one location in such a way that it can be "reconstructed [at another] from purely classical information and pure nonclassical Einstein-Podolsky-Rosen (EPR) correlations[3]". Reference to 'EPR correlations' means that Alice, the sender, and Bob, the receiver, each have in store one particle from a suitable **A+B** pair, such that the $\mathbf{H}_A \otimes \mathbf{H}_B$ ket that is ascribed to the **A+B** preparation is not factorable into $\mathbf{H}_A$ and $\mathbf{H}_B$ kets. If **A** and **B** are two *qubits*[4] e.g. two spin-½ particles, an appropriate ket will be for example

$$|\Psi^-\rangle_{AB} = 2^{-\frac{1}{2}}(|01\rangle - |10\rangle), \qquad (8.1.1)$$

where $|0\rangle$ and $|1\rangle$ are QIT (Dirac) notations for eigenkets of σ_z, such that $\sigma_z|0\rangle = |0\rangle$ and $\sigma_z|1\rangle = -|1\rangle$. (8.1.1) is characteristic of a so-called *singlet*[5].

A third party, Claire, supplies Alice with another qubit **C**. In the $\{|0\rangle, |1\rangle\}$ basis, the ket ascribed to that qubit is of the form $|X\rangle_C = a|0\rangle + b|1\rangle$, but the values of *a* and *b* are not disclosed by Claire. Obviously, if Alice knew everything about the way Claire prepared **C**, she could just inform Bob about

---

[1] Bennett *et al.* 1993.
[2] *Ibid.*
[3] *Ibid.*
[4] The oddly spelt word *qubit* (pronounced 'kewbit'), coined by B. Schumacher (1995), is just another name for a two-level system. 'Qubit' is actually a misnomer, conveying as it does the idea of a unit of information that would be specific to quantum theory. Claiming that "there is a lot more than one bit of information in a qubit" (Deutsch and Hayden 1999) is just another way of pointing out the 'intermediacy' (Dirac 1958) of linear combinations of basis kets $|0\rangle$ and $|1\rangle$, which is the key to devising secure cryptosystems or fast algorithms. these feats exploit the characteristic asymmetry mentioned in Section 6.1: whereas there is an infinite number of possible qubit preparations, the amount of information that can be gained empirically from a qubit is just 1 bit (all of the relevant observables are two-valued).
[5] This is such that a measurement of any given spin component on **A** and **B** (the same each) yields opposite values (strict anticorrelation).

$|X>_C$ and that information would allow him to prepare a particle of his own 'in the same state'. If, on the other hand, Alice is not privy to Claire's preparation, this trivial option is not available, and this is precisely what makes the claim that Bob can still recover $|X>$ so intriguing.

Measuring a two-valued observable like σ_x or σ_z on a single qubit supplies only 1 bit (there are two possible outcomes). This is greatly insufficient to determine both *a* and *b*. As a matter of fact, no measurement performed on **C** could provide Alice with the information she needs to identify $|X>_C$. Another solution that comes to mind is that, using identically prepared particles, she could perhaps produce many copies of $|X>_C$ and then work out a density matrix, using the mean values obtained in measuring appropriate observables on those copies. The 'no-cloning' theorem[1], however, precludes any general (unitary) procedure whereby copies of quantum states could be generated at will, without having first to identify those states. None the less, by taking advantage of the entanglement of **A** and **B** Alice can still achieve what seems *prima facie* to be impossible. A simple protocol involving both $|X>_C$ and $|\Psi^-\rangle_{AB}$ makes it possible for Bob to 'recover' $|X>$ as the 'state' of his own qubit, whatever $|X>$ may be.

The initial 3-qubit product

$$|\Psi_i> = |X>_C \otimes |\Psi^->_{AB} = 2^{-1/2}(a|0>_C + b|1>_C)(|01>_{AB} - |10>_{AB}) , \qquad (8.1.2)$$

can be expanded as

---

[1] The *no-cloning theorem* (Wootters and Zurek 1982, Dieks 1982) is a straightforward consequence of linearity in the SAQM. As the word 'cloning' suggests, the idea is to end up with two copies of the *same* 'state' $|u>$ as a result of a linear, unitary transformation **V**, which takes an initial product $|u>\otimes|s> \in \mathbf{H}_1\otimes\mathbf{H}_2$ into $|u>\otimes|u>$, for any input $|u>$. If **V** exists, then $\mathbf{V}[a|u>+b|v>)\otimes|s>]$ should be both equal to $(a|u>+b|v>)\otimes(a|u>+b|v>)$ and to $a\mathbf{V}(|u>\otimes|s>) + b\mathbf{V}(|v>\otimes|s>) = a|u>\otimes|u> + b|v>\otimes|v>$. This is clearly inconsistent if neither a nor b vanish. Notice that, since **V** is unitary ($\mathbf{V}^\dagger\mathbf{V}=\mathbf{I}\otimes\mathbf{I}$), then $<u;s|\mathbf{V}^\dagger\mathbf{V}|v;s>$ must be equal both to $<u;u|v;v> = <u|v>^2$ and to $<u|v>$. If $|u>$ and $|v>$ are not identical, these two results agree only if the vectors are mutually orthogonal: members of a set of clonable 'states' must be orthogonal to one another. This theorem was later found to be a special case of another (*no-broadcasting theorem*, Barnum *et al.* 1996.), which is applicable to both the mixed and the pure case. 'Broadcasting' here means a $\mathbf{H}_1\otimes\mathbf{H}_2$ transformation which, given an input statistical operator ρ acting upon $\mathbf{H}_1$, yields marginal statistical operators that are empirically indistinguishable from ρ (in other words, the transformation is $\rho\otimes\sigma \rightarrow \rho'$, such that $tr_1(\rho') = tr_2(\rho') = \rho$). Pure case broadcasting is clearly just cloning. However, broadcasting is not simply mixed state cloning, for ρ' is not required to be factorable. The theorem says that broadcasting can be achieved for members of a given set of statistical operators if and only if the operators in that set commute. No-broadcasting and no-cloning are actually consequences of a more general mathematical fact: completely positive maps, which encompass all lawful forms of time evolution in quantum theory, cannot increase distinguishability (in the sense of Section 6.2); Fuchs and van de Graaf 1999; Nielsen and Chuang 2000.

$$|\Psi_i\rangle = 2^{-1/2}(a|001\rangle - a|010\rangle + b|101\rangle - b|110\rangle), \qquad (8.1.3)$$

where the first two 'bit slots' in every three-place ket refer, in this order, to Alice's qubits **C** and **A**, the third to Bob's. More significantly, (8.1.2) can be written in terms of the four normalised and mutually orthogonal 'Bell' kets $|\Psi^{\pm}\rangle = 2^{-1/2}(|01\rangle \pm |10\rangle)$ and $|\Phi^{\pm}\rangle = 2^{-1/2}(|00\rangle \pm |11\rangle)$:

$$|\Psi_i\rangle = \frac{1}{2}(|\Phi^+\rangle_{CA}|u_3\rangle_B + |\Phi^-\rangle|u_2\rangle - |\Psi^+\rangle|u_1\rangle - |\Psi^-\rangle|u_0\rangle), \qquad (8.1.4)$$

where $|u_0\rangle = |X\rangle$, $|u_1\rangle = \sigma_z|X\rangle$, $|u_2\rangle = \sigma_x|X\rangle$ and $|u_3\rangle = \sigma_x\sigma_z|X\rangle$.

It is easily seen how what has been promised can now be delivered. Alice performs on *both* **C** *and* **A** a 'Bell analysis', which amounts to a discriminating operation among the four Bell 'states'[1]. As a result of the interaction of **C** and **A** within an appropriate device (Bell analyser) and regardless of the input $|X\rangle_C$, one of four equally probable outcomes will randomly obtain. Suppose that the 'Bell measurement' picks out $|\Phi^-\rangle$. Since letting **A** and **C** immediately through the analyser again would yield with certainty the same outcome, that measurement can be conveniently thought of as effecting, symbolically, a projection of the initial ket $|\Psi_i\rangle$ onto one of the four products that are linearly combined in (8.1.4), namely $|\Xi\rangle = |\Phi^-\rangle|u_2\rangle$. A definite outcome of the Bell measurement provides an unambiguous answer to a 2-bit question. Alice sends those 2 bits to Bob, whereby he learns how the 'state' of his qubit *differs* from that of Claire's, whatever that state was. This follows from the one-to-one correspondence, explicit in (8.1.4), between each of the Bell kets and the result of a simple unitary transform $|u_i\rangle$ of $|X\rangle$. In the example, Bob just has to perform a NOT operation (σ_x) to 'recover' the state. Depending on the outcome of Alice's Bell measurement, Bob may have to perform a phase shift (σ_z), a NOT followed by a phase shift ($\sigma_z\sigma_x$), or indeed do nothing at all if Alice informs him that she picked out a singlet.

After Alice's measurement has been performed[2] and assuming a symbolic reduction to $|\Xi\rangle$ of the initial ket $|\Psi_i\rangle$, the marginal statistical operator associated with Claire's qubit is

---

[1] For a discussion of the explicit form of the 'Bell observable', see Braunstein *et al.* 1992.

[2] Claire's *particle* is not required to be destroyed in the process. In familiar science fiction accounts of 'teleportation', an object or person is said to be 'disassembled' at one place to be 'reassembled' somewhere else; the teleportation device is usually portrayed as a kind of sophisticated 3-D fax machine that scans-and-destroys the original. Although such destruction is not necessary, it is generally assumed in order to avoid 'who is I' conundrums that would arise if a copy of a 'teleported' person were to coexist with the original (see Parfit

$$\rho_C^f = tr_{AB}(|\Xi><\Xi|) = \frac{1}{2}(|0><0| + |1><1|) = \frac{1}{2}I_C, \qquad (8.1.5)$$

where I_C is the identity operator with respect to the Hilbert space $\mathbf{H}_C$. (8.1.5) is sometimes said to correspond to a *maximally mixed state*. The irreversible transformation of $\rho_C^i = |X><X|_C$ into ρ_C^f is usually taken as evidence that quantum teleportation complies with the no-cloning theorem. However, since no unitary transform can possibly map a linear combination like (8.1.4) into a ket product, the no-cloning theorem is not strictly speaking applicable to the transformation described above. Besides, if any 'loss' can be said to occur in the process of Bell analysis, it is only with respect to the predictions Alice *could* have made using the ket $|X>_C$, had she known what it was from the outset. Since that knowledge was not available to her in the first place, she is not better off (nor worse) in that respect afterwards.

It is essential for the success of the whole procedure that Alice and Bob 'share entanglement[1]'. Suppose that the **A+B** pair has been prepared according to, say, $|01>_{AB}$ (performing a measurement of $\sigma_z^{(A)}$ would yield with certainty +1 whereas a measurement of $\sigma_z^{(B)}$ would yield -1). The initial **C+A+B** ket would then be

$$|X>\otimes|0>\otimes|1> = a|001> + b|101>$$
$$= \tfrac{1}{2}(a|\Phi^+> + a|\Phi^-> + b|\Psi^+> - b|\Psi^->)|1>: \qquad (8.1.6)$$

the ket ascribed to **B** would remain $|1>$ whatever the outcome of Alice's Bell measurement. Since **A** and **B** are not correlated, no teleportation can be achieved (and no 2-bit message sent by Alice would be of any help to Bob). The same conclusion holds if Alice and Bob's pairs originate from an ensemble associated with a statistical operator like

$$\rho_{AB} = \frac{1}{2}(|01><01| + |10><10|). \qquad (8.1.7)$$

Such a mixture can be (non-uniquely) realised by drawing **A+B** pairs at random from a collection in which 50% of those pairs are correctly 'described' using $|01>$, 50% using $|10>$. **A** and **B** spins are then correlated: spin components of the members of the same pair in the z direction are always found opposite. To **C+A+B** is then associated the diagonal statistical operator $\rho_{CAB} = |X><X|_C \otimes \rho_{AB}$. There again, the absence of entanglement

---

1987). The discussion should make it clear that quantum teleportation hardly has anything to do with any sort of fax, be it based on quantum-mechanical principles.

[1] It is sometimes said that Alice and Bob 'share 1 *ebit*'. The word *ebit* ('e' for *entanglement*) is used to denote a single instance of maximal entanglement, i.e. a binary preparation associated with a Bell ket. The remarks made about the word 'qubit' apply, with the difference that the maximum empirical yield of an ebit is just 2 bits.

makes it impossible to achieve teleportation[1]: after the Bell analysis, $\rho_B^f = tr_{CA}(\rho_{CAB}) = \frac{1}{2} I_B$, which is different from $|X><X|$ and from any unitary transform thereof.

No less crucial to teleportation are Alice's 2 bits. Failing to be informed of her result, Bob could not do any better than bet on one of the four equally likely possibilities exhibited in (8.1.4). Acting on **B** accordingly, he would then hopefully, but with a 75% chance of being proven wrong later by Claire, prepare his own qubit 'as' $|X>_B$. All he could rightly claim is that the final 'state' of **B** is related to $|X>$ by *some* transformation σ_x, σ_z, $\sigma_x\sigma_z$ or I_B (together, these transformations form a group **P**). Besides, that the 2 bits must be 'classically' conveyed e.g. broadcast or transmitted over the phone, also ensures that no observable conflict with STR occurs: it cannot take less time for Alice's message to reach Bob than what it takes for a signal to propagate between their two locations at the maximum speed **c**. This rules out any superluminal signalling based on iterating the procedure.

As long as Bob cannot compare lists of his own measurement results with Alice's, the marginal statistical operator ρ_B is the only predictive handle he has on his own particle(s). Before Alice Bell-measures **C+A**,

$$\rho_B^i = tr_{CA}(|\Psi_i><\Psi_i|) = \frac{1}{2}(|0><0| + |1><1|) = \frac{1}{2} I_B \qquad (8.1.8)$$

Were Bob to measure $\sigma_z^{(B)}$, he would therefore get either +1 or -1 with equal probability. Since this operator can also be written $\rho_B^i = \frac{1}{4}\sum_{k=0}^{k=3} |u_k\rangle\langle u_k|$ then, until he receives Alice's message, and regardless of whether she had by then acted at all on her qubits, he might as well believe that **B** is 'already' characterised by one of the four kets $|u_i>$, except that he does not – cannot – know which. His ignorance is maximal and it takes the 2 bits Alice supplies for him to be able to 'recover the right state'.

Because of the non-factorability of (8.1.4), probabilities of predicted outcomes at Bob's end are 'strongly dependent' on the outcomes of measurements carried out by Alice, in the following sense: whilst the probability that Alice obtains $|\Phi^->_{CA}$ and that Bob 'ends up with' $|u_2>_B$ is $p_{\Psi_i}(\Phi^-_{CA}\ \&\ u_{2,B}) = ¼$, the probability that the Bell measurement yields $|\Phi^->_{CA}$ is $p_{\Psi_i}(\Phi^-_{CA}) = ¼$, which implies that the conditional probability $p_{\Psi i}(u_{2,B} / \Phi^-_{CA}) = 1$. However, the value 1 of that probability does *not* imply that the outcome associated with $|u_2>_B$ is itself determined by the actual, recorded occurrence of the Bell measurement result associated with $|\Phi^->_{CA}$.

---

[1] Popescu 1994.

Suppose that Claire provides Alice with a large number of particles, all identically prepared according to $|X\rangle_C$. Instead of patiently waiting for Alice's information, Bob measures $\sigma_z^{(B)}$ on some of his own particles. If he gets, say, +1 for the **B** particle of pair No. 326 at time t_B, then he can legitimately assign to $\mathbf{C+(A+B)}_{326}$

$$|\Lambda\rangle = \tfrac{1}{2}\,(a|\Psi^+\rangle + a|\Psi^-\rangle + b|\Phi^+\rangle - b|\Phi^-\rangle)|0\rangle \qquad (8.1.9)$$

(his assignment is (8.1.6) if the result is -1). Unaware of Bob's doings, Alice subsequently measures $\mathbf{C+A}_{326}$ and obtains a result that corresponds to one of the four Bell kets. That result is just as compatible with (8.1.9) or (8.1.6) as it is with (8.1.4). There is no way she can infer from the outcome what the 'right' ket for $\mathbf{C+(A+B)}_{326}$ was (if indeed 'right' means anything in this context) immediately before the time $t_A > t_B$ when she performed her measurement, i.e. distinguish, using that measurement, between (8.1.4) and (8.1.9).

The adherent to an ontic, single-case view of quantum states might object that, whatever the outcome of Alice's Bell measurement, the state of $\mathbf{B}_{326}$ did physically 'collapse' at t_B into the eigenstate $|0\rangle$ of $\sigma_z^{(B)}$ as a result of Bob's (previous) measurement. There would thus be a fact of the matter regarding (8.1.9), rather than (8.1.4), being the 'right' **C+A+B**-state immediately before t_A. If so however, since $\mathbf{C+A}_{326}$ in (8.1.9) is uncorrelated to **B**, the state of $\mathbf{B}_{326}$ is none[1] of the $|u_i\rangle$ at the time $t_B' > t_A > t_B$ when Bob is informed of the result of Alice's Bell measurement on $\mathbf{C+A}_{326}$. As a result, regardless of Alice's instructions, none of the four unitary transformations associated with the Bell-measurement outcomes will 'put' $\mathbf{B}_{326}$ 'in' the intended state $|X\rangle_C$. The problem with such reasoning is that it overlooks the fundamentally statistical significance of the quantities and operations involved. No measurement performed on $\mathbf{B}_{326}$ alone will make it possible to determine 'its state'. On the other hand, if Claire supplies Alice with many identically prepared **C** particles, Bob can later compare his own table of $\sigma_z^{(B)}$ outcomes with the outcomes of Alice's Bell measurements. Matching **A**'s and **B**'s from the same pairs, he can select in his table those entries for which Alice's measurement corresponds to a given Bell ket, e.g. $|\Phi^+\rangle$. Provided the list is large enough, his working out the statistics of his own measurement outcomes as recorded on the **B**-list will convince him they are compatible with

[1] Unless, perhaps, $|X\rangle_C$ happens to be an eigenstate of $\sigma_z^{(C)}$ – but this neither Alice nor Bob knows anyway.

ascribing the ket $|X>_B$, up to a **P**-transformation, to the corresponding **B**-ensemble (compare (8.1.6) and (8.1.9)): for a sufficiently large Φ^+-list, the frequency of '0' outcomes among his $\sigma_z^{(B)}$ measurements will, in agreement with (8.1.9), approximate the probability $|b|^2$. But this is just the probability distribution that will obtain if, instead of performing any measurement prior to Alice's, Bob measures $\sigma_z^{(B)}$ *after* t_A and then works out the statistics of the results for those of his qubits that are associated with a $|\Phi^+>$ outcome. Bob will, of course, not be able to make any such selection unless Alice provides him with the relevant information, i.e. with her own list of measurement outcomes (2 bits per item).

Bell analysis hinges on a very special kind of interaction of **C** and **A**. However, the required interaction can be dissociated from measurement without changing the essentials of 'teleportation'. Instead of a Bell measurement on **C+A**, Alice may choose to carry out a *controlled not* (**CNOT**) operation on **C** and **A**, where 'controlled' means that the value of the **C**-bit in a $\mathbf{H}_C \otimes \mathbf{H}_A$ product of kets determines whether or not a change occurs in the value of the **A**-bit. More precisely, a bit flip (**NOT**) is applied to the target **A** bit if and only if the value of the control **C** bit is 1:

$$\mathbf{CNOT}|j, k> = |j, j \oplus k> \quad (j,k = 0,1),$$

where $\oplus$ denotes the addition modulo 2. Two-qubit basis kets $|00>$, $|01>$, $|10>$, $|11>$ are therefore mapped into $|00>$, $|01>$, $|11>$ and $|10>$ respectively. This operation amounts to an XOR (exclusive OR) being performed on the two bits and coding the result into the second bit. Applied to (8.1.2), this yields

$$\begin{aligned} |\Delta> &= (\mathbf{CNOT}_{CA} \otimes \mathbf{I}_B)|\Psi_i> \\ &= 2^{-1/2}(a|00>|1> - a|01>|0> + b|11>|1> - b|10>|0>) \end{aligned} \tag{8.1.10}$$

or

$$|\Delta> = \frac{1}{2}[|0'0>|u_3> + |1'0>|u_2> - |0'1>|u_1> - |1'1>|u_0>, \tag{8.1.11}$$

where $|0'> = 2^{-1/2}(|0> + |1>)$ and $|1'> = 2^{-1/2}(|0> - |1>)$ are eigenkets of σ_x ('NOT gate') associated with the eigenvalues +1 and -1 respectively.

As a result of the **CNOT** interaction, the marginal statistical operator for Claire's qubit becomes $\rho_C^{(\Delta)} = |a|^2 |0><0| + |b|^2 |1><1| \neq \rho_C^i$. The two operators cannot be distinguished on the basis of measuring $\sigma_z^{(C)}$ on the single particle **C**, for the outcome would be +1 or -1 in either case. On the

other hand, measuring $\sigma_x^{(C)}$ *and* $\sigma_z^{(A)}$ provides Alice with the two bits required for Bob to recover the input state[1] (with the above qualifications regarding the meaning of that 'recovery'). If $\mathbf{CNOT}_{CA}$ is followed by a Hadamard transform[2] $\mathbf{H}^{(C)}$ on **C** (the final **C+A+B** ket $|\Gamma>$ has the same form as (8.1.11) with 0 replacing 0', 1 replacing 1'), a measurement of $\sigma_z^{(C)}\otimes\sigma_z^{(A)}$ would actually suffice. The only 'cost' in involving just one observable explicitly[3] is that the $\mathbf{H}_C$-related marginal statistical operator becomes maximally mixed: $\rho_C^{(\Gamma)} = \frac{1}{2}\mathbf{I}_C$ (this means that, if Alice were to repeat the "$\mathbf{CNOT}_{CA}$, then $\mathbf{H}^{(C)}$" sequence on identically prepared **C**'s and **A**'s, subsequent measurements of any spin component on **C** would yield +1 and -1 with equal probability). The $\rho_C^{i} \rightarrow \rho_C^{(\Gamma)}$ transformation can be reversed by applying the sequence twice. Therefore, if an irreversible change in the marginal statistical ρ_C is taken as a criterion for teleportation to be achieved, the latter is not completed until Alice records a definite outcome and applies the reduction rule (**P7'**) in order to update her **C+A+B** ket assignment.

8.2. The fastest way from A to B?

It is tempting to think of Alice's intervention as resulting in a physical change in the **B** particle, a change that is made possible by the 'EPR link' provided by the **A+B** entanglement. The question then arises of how local interactions of **C** and **A** can possibly affect Bob's qubit at a distance. Even if we admit that "some sort of action-at-a-distance or (conceptually distinct) nonseparability seems built into any reasonable attempt to understand the quantum view of

---

[1] The four recovery (**P**) operations remain the same as they were in the original scheme.

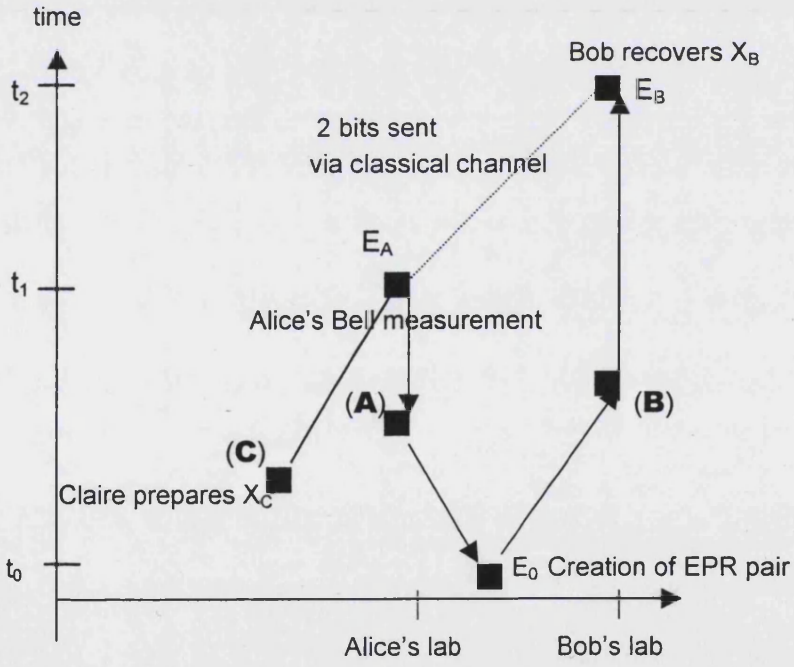
[2] The *Hadamard* transform **H**, such that (up to normalisation) $\mathbf{H}|k> = |0> + (-1)^k|1>$ with k = 0 or 1, amounts to a reflection in an axis whose direction is at an angle $\frac{\pi}{8}$ from the $|0>$ axis. The corresponding matrix is $\mathbf{H} = 2^{-1/2}(\sigma_x + \sigma_z) = 2^{-1/2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, which is its own inverse: $\mathbf{H}^2 = \mathbf{I}$. A measurement of σ_z after application of **H** to either basis ket $|0>$ or $|1>$ yields either bit value with equal probability: **H** yields a 'random bit'. The different relative phases exhibited by $\mathbf{H}|0> = |0'>$ and $\mathbf{H}|1> = |1'>$ are exploited for the reinforcement or the cancellation of terms in an automatic computation. That **H** is a Fourier transform on the additive group of integers modulo 2 is also essential to the implementation e.g. of Shor's fast factoring algorithm (Shor 1994).

[3] Reference to another observable (σ_x) remains implicit since the Hadamard transform maps the $\{|0>, |1>\}$ basis into the dual basis $\{|0'>, |1'>\}$ of σ_x eigenvectors.

reality[1]", most physicists are reluctant to contemplate the possibility of what Einstein dismissed as unacceptable spooky actions at a distance (*Spukhafte Fernwirkungen*). One way of retaining the idea of a 'teletransport' without introducing spooky influences is to allow backwards-in-time propagation of an essential part of what is deemed transferred. In the words of Richard Jozsa[2],

> Alice succeeds in transferring the quantum state $|\psi>$ to Bob by sending him just two bits of classical information. Clearly these two bits are vastly inadequate to specify the state $|\psi>$ so how does the remaining information get across to Bob? What carries it? What route does it take?... there is clearly only one other route connecting Alice to Bob (apart from the channel carrying the two classical bits) – it runs *backwards in time* from Alice to the creation of the EPR pair and then forwards in time to Bob. Hence we must conclude that most of the quantum information of $|\psi>$ was propagated along this route, firstly backwards in time and then forwards to Bob.

This proposal, endorsed by prominent figures in QIT like Charles Bennett or Sandu Popescu, is summarised in the space-time diagram below. The arrows indicate the direction of flow of what Jozsa refers to as quantum information.



"With the above routing, we can deduce", according to Jozsa, "that at times between t_0 [creation of the EPR pair] and t_1 [Alice's measurement] most of the quantum information was already [?] well on its way to Bob (in his arm of the EPR pair) even though Alice had not yet performed her measurement.

---

[1] Redhead 1987, p.169; see Mermin 1999 and Fuchs 2001 for dissenting views about the relevance of non-locality to quantum theory.
[2] in Lo *et al.* 1998, p.58.

Indeed she may not yet have decided to transmit anything to Bob or even have been born yet![1]". Roger Penrose, who subscribes to Jozsa's picture, claims that an "especially noteworthy[2]" aspect of quantum information is that it "is not constrained by the usual spatiotemporal 'causality' requirements of relativity[3]". Whilst entanglement cannot be used to enhance signalling beyond the relativistic (speed-of-light) limit, there is "nothing against quantum information 'travelling backwards in time'; indeed, such curious behaviour seems to be *required*[4] "...

Jozsa's answer to the question "Quantum Information – What is it[5]?" is hardly explanatory: it is information that "is embodied in a given unknown quantum state[6]". Although it is mostly "inaccessible", it is "still useful – for example it is necessary in its totality to correctly predict any future evolution of the state and to carry out the processes of quantum computation[7]." This is pretty much all that QIT theorists usually have to say about quantum information, and is at best another way of alluding to the specific traits of a Hilbert space-based predictive scheme, and to new possibilities in computing, cryptography etc. the implementation of such a scheme affords. Understandably, QIT workers revel in exploiting the representative capacities of Hilbert space to computational or cryptographic ends. It is also remarkable that we should be able at all to recover or reconstruct the full statistical make-up of a preparation about which no information, as classically conceived, is actually available. However, do such feats (which, to this day, remain more theoretical than effective[8]) require introducing a new kind of information? If the calculation of probabilistic connections between inputs ('state' assignments) and selected outputs (values associated with measurement outcomes) calls for a drastic departure from an ordinary version of the probability calculus, this by itself does not warrant the claim that some specifically 'quantal' form of information comes into play (no more, indeed, than that the 'kolmogorovity' of probability is undermined). Penrose and Jozsa's invocation of quantum information does little more than articulate their intuition that since 'teleporting an unknown (quantum) state' is classically impossible, certainly 'more than just 2 bits' are required. Those 2 bits would be needed only to ensure that Bob recovers *the*

---

[1] *Ibid.*

[2] Penrose 1998, p.1928. Penrose's view of quantum information ("an aspect of quantum entanglement") appears to be more restrictive than Jozsa's.

[3] Penrose 1998, p.1928.

[4] *Ibid.* (Penrose's emphasis).

[5] *Ibid.*, p.49.

[6] *Ibid.*

[7] *Ibid.*,p.50.

[8] Advances in nonlinear optics have made it possible, at least, to test the feasibility of quantum teleportation with polarised photons; all of the tests confirm the predictions of quantum theory (Mattle *et al.* 1996; Boschi *et al.* 1998; Kim *et al.* 2001). On the other hand, Shor's factoring algorithm is unlikely to run on a quantum computer in the near future.

intended state, rather than some **P**-transform thereof. 'Something else', then, must have travelled or been exchanged one way or another between parts of the **A+B+C** system to guarantee that the final state of Bob's qubit is, up to a **P**-transformation, identical with the initial state of Claire's qubit. If a superluminal link is rejected, then there is little choice but to conceive of some kind of backwards-in-time information transfer from the **A** particle to the **A+B** pair, along the space-time path that links the two events E_A and E_0, 'followed' by a forward propagation of that information up to the time when Bob acts upon **B** according to Alice's instructions (E_B). Penrose and Jozsa never point out, however, how radical a shift in our understanding of information such an option calls for. Quantum information would have little if anything to do with the precise notion variously formalised since R. Fisher and C. Shannon. Whereas information in Shannon's sense[1] comes down to a simple logarithmic measure of average 'surprisal[2]', quantum information would have to be an objective feature of the physical world. Being generated as a result (?) of physical interactions, this novel addition to the category of substance would propagate – forward and possibly backward in time, if the latter makes any good physical sense – and determine at a distance some attributes of physical systems...Needless to say, there is not a shred of evidence in support of the existence of any form of backwards-in-time propagation in the physical world. Besides, as Jozsa himself concedes, "for all physical[3] purposes, in Bob's laboratory his original EPR particle is completely indistinguishable from an equal probabilistic mixture of the four states. Thus until the two bits of classical information arrive from Alice, the quantum information (which flowed partly backwards in time) is *utterly useless* and *completely indistinguishable* from the situation in which Alice performed no measurement at all[4]." The "inherent inaccessibility[5]" of that information hardly speaks in favour of the proposal. However, Jozsa argues that it is precisely because the two mixtures $\frac{1}{4}\sum_{k=0}^{k=3}|u_k\rangle\langle u_k|$ and $tr_{CA}(|\Psi_i><\Psi_i|)$, both equal to $\frac{1}{2}I_B$, cannot be empirically distinguished that one might just as well claim, without any threat to causality, that the 'state' of Bob's particle *was,* between t_0 and t_1, already (?) a mixture of the four $|u_k>$. Should we understand that Bob's qubit was actually 'in' one of the $|u_k>$, but that Bob just ignored which one – and could not possibly find out? But since it is improper (being obtained by a partial tracing operation), the

---

[1] Shannon 1948.

[2] The word 'surprisal', which is evocative of an observer's response to the occurrence of a given symbol, was coined by Tribus 1961 (Shannon's information equals the average surprisal in a message only in the absence of noise) ; see also Dretske 1981.

[3] By what Jozsa seems to mean *practical.*

[4] Jozsa, in Lo *et al.* 1998, p.60 (italics mine).

[5] *Ibid.*

mixture described by $\frac{1}{4}\sum_{k=0}^{k=3}|u_k\rangle\langle u_k|$ *cannot* be regarded as an ensemble of **B** particles, each of which would definitely be *in* one of the $|u_k\rangle$ – and why in *those* $|u_k\rangle$ for that matter rather than, say, some eigenstate of σ_z?

Suppose that we decide not to dismiss altogether the suggestion of quantum information flow, be it partly backwards in time. Are the interactions involved in Bell analysis or the CNOT responsible for its being released? Or is backwards-in-time propagation triggered by a 'collapse' induced by Alice's acting upon both **C** and **A**? In the CNOT picture, is the release delayed until Alice makes her spin measurements and obtains definite results, or – less plausibly still – does it occur in a piecewise fashion, partly as a result of the CNOT interaction, partly as a consequence of collapse? Let us modify the standard procedure so that it now involves, instead of just Alice and Bob, a number of intermediate agents. Nothing prevents in principle n = 2p qubits to be prepared as

$$|\Psi_n\rangle = 2^{-1/2}(|i_1\, i_2 \ldots i_n\rangle + |j_1\, j_2 \ldots j_n\rangle), \qquad (8.2.1)$$

where for all k = 1 to n, i_k = 0 or 1 and $j_k = 1 - i_k$, hence $\langle i_1\, i_2 \ldots i_N | j_1\, j_2 \ldots j_N\rangle = 0$. These n qubits $\mathbf{Q}_i$ may be distributed over p+1 agents so that Bob_1 receives $\mathbf{Q}_1$, Bob_2 receives $\mathbf{Q}_2$ and $\mathbf{Q}_3$, ... Bob_p receives $\mathbf{Q}_{n-2}$ and $\mathbf{Q}_{n-1}$ and the last in the chain, Bob_{p+1}, receives $\mathbf{Q}_n$. Claire supplies Bob_1 with a qubit **C** whose 'state' X is unknown to all of the Bobs. The ket ascribed to the (n+1)-qubit system might be, for example,

$$|\Sigma_0\rangle = |X\rangle_C|\Psi_n\rangle = 2^{-1/2}(a|0\rangle + b|1\rangle)(|0101\ldots01\rangle + |1010\ldots10\rangle), \qquad (8.2.2)$$

which can also be written

$$|\Sigma_0\rangle = \frac{1}{2}[|\Psi^+\rangle(a|01\ldots0\rangle + b|10\ldots1\rangle) + |\Psi^-\rangle(a|01\ldots0\rangle - b|10\ldots1\rangle) + |\Phi^+\rangle(a|10\ldots1\rangle + b|01\ldots0\rangle) + |\Phi^-\rangle(a|10\ldots1\rangle - b|01\ldots0\rangle)]. \qquad (8.2.3)$$

Let n= 6. At time t_1, Bob_1 performs a Bell measurement of $\mathbf{C}+\mathbf{Q}_1$. If the result corresponds, say, to $|\Phi^+\rangle_{C1}$, Bob_1 can update the initial ket to

$$\begin{aligned}|\Sigma_1\rangle &= |\Phi^+\rangle_{C1}(a|10101\rangle + b|01010\rangle) \\ &= 2^{-1/2}|\Phi^+_{C1}\rangle[|\Psi^+\rangle_{23}(a|101\rangle + b|010\rangle)_{456} \\ &\quad + |\Psi^-\rangle_{23}(-a|101\rangle + b|010\rangle)_{456}\end{aligned} \qquad (8.2.4)$$

Without being informed of the result, Bob_2 subsequently (at t_2) Bell-measures $\mathbf{Q}_2+\mathbf{Q}_3$. Assuming that the outcome is that which corresponds to $|\Psi^- >_{23}$, whoever knows the outcome of Bob_2's and Bob_1's measurement will update (8.2.4) to

$$|\Sigma_2> = |\Phi^+>_{C1}|\Psi^->_{23}(-a|101>_{456} + b|010>)$$
$$= 2^{-1/2}|\Phi^+>_{C1}|\Psi^->_{23}(|\Psi^+>_{45}|u_3>_6 + |\Psi^->|u_2>). \qquad (8.2.5)$$

Eventually, Bob_3 measures $\mathbf{Q}_4+\mathbf{Q}_5$ at t_3 and he finds, say, $|\Psi^->_{45}$. In order to 'recover' $|X>_C$, Bob_4 must first receive 2 bits from Bob_1, 1 bit from Bob_2 and 1 bit from Bob_3, allowing him to write the 'final' ket as

$$|\Sigma_f> = |\Phi^+>_{C1}|\Psi^->_{23}|\Psi^->_{45}|u_2>. \qquad (8.2.6)$$

The cost of the recovery cannot be less than 4 bits, or more generally p+1 bits. Bob_4 then knows which **P** transformation he must apply (σ_x in the example) to complete this variant of teleportation. If the process is supposed to involve any quantum information transfer from $\mathbf{Q}_1$ to $\mathbf{Q}_n$, then this transfer appears to be achieved piecewise rather than all-at-once (any of the Bobs may choose to postpone his own intervention as long as he pleases). This would appear to require a fine tuning of the 'partial quantum information' released at different times t_1, t_2, t_3 to properly affect $\mathbf{Q}_1 +...+ \mathbf{Q}_n$ prior to the first Bob's intervention and yield the 'right' teleported state, casting more doubt on the plausibility of the Jozsa-Penrose account. This can, however, not really threaten a proposal that does not lend itself to experimental refutation.

According to Maroney and Hiley, the Bohm-Hiley interpretation would be unique in affording a clear physical picture of what goes on in quantum teleportation. They aim to show, in particular, that loose invocations of 'quantum information' and its alleged transfer can be advantageously replaced with an explanation in terms of the transfer of objective 'active information'. Bearing in mind remarks made in Section 3.2, the expression *BH-activity* is used in what follows in place of Bohm, Hiley and Maroney's 'active information'. Likewise, *BH-passivity* is used instead of their 'passive information'[1] (by what is simply meant lack of active information).
If a system ('particle') is held to be 'described' using a superposition of two non-overlapping wave functions or packets, the Bohm-Hiley interpretation has it that the objective position of the particle will be located within either one or the other of the packets. The kinematical behaviour of the particle will then only depend on the action of the quantum potential on the wave packet that does 'contain' the particle or, to put it another way, the resulting subquantal trajectory will then only depend upon the mathematical form of that wave

[1] Bohm and Hiley 1993.

packet. That particular wave packet is said to be endowed with BH-activity, whereas the other one is BH-passive. Inducing the overlap of the two wave packets through some appropriate interaction will have the effect of activating the previously BH-passive packet. The surge of BH-activity that is supposed to be prompted by the recombination will then affect the particle's trajectory via what, at the wave function level of description, corresponds to complex interference effects.

Now consider the case of a two-particle entangled wave function of the positions x_A, x_B of the particles A and B:

$$\Psi(x_A, x_B) = \phi_\alpha(x_A)\chi_\alpha(x_B) + \phi_\beta(x_A)\chi_\beta(x_B) \qquad (8.2.7)$$

In accordance with the 'common pool of information' idea, BH-activity will then depend upon the simultaneous position of both particles. Assuming that the ϕ wave functions are overlapping but that the χ functions are not and that, say, the objective position of the A particle lies within ϕ_α only, then the only BH-active contribution to $\Psi(x_A, x_B)$ will be $\phi_\alpha(x_A)\chi_\alpha(x_B)$, whereas the other 'branch' will be BH-passive. As a result, particle A will be affected by ϕ_α only, despite the ϕ overlap. If the latter is physically suppressed, particle B will only ever be found within $\chi_\alpha(x_B)$. If the ϕ overlap is restored afterwards, BH-activity will remain confined within the 'α part', because of the absence of χ overlap. BH-activating the BH-passive branch of the superposition would require that ϕ and χ functions in the α and β branches be simultaneously brought back into overlapping positions, which is generally hardly achievable if not impossible[1].

Spin is accounted for by including in the polar decomposition of the total wave function a spin factor, which has the form of a spinor with unit magnitude and zero average phase. The resulting modified Hamilton-Jacobi-like equation then includes a spin quantum potential, which determines through a 'quantum torque[2]' the BH-activation of the spinor wave function, hence the evolution of the spinor states, over and above the effect of the classical interaction of the magnetic dipole moment of the particle with an external magnetic field.

Maroney and Hiley assume that each of the three particles A,B,C involved in the teleportation protocol can be ascribed a localised position wave packet. Initially, the three wave packets are not entangled: the trajectory of each particle is determined separately by a BH-activating quantum potential. On the other hand, the spin degrees of freedom of A and B are linked via the spin quantum potential. If a spin measurement is performed on particle A alone, the

---

[1] The 'measurement problem' is solved or, rather, does not arise on that account since, if one particle is replaced with a macroscopic measuring device, it is deemed impossible for the restoration of BH-activity to be achievable via any controllable interaction. The resulting permanent de-BH-activation of one or more branches plays in the Bohm-Hiley approach the part of wave function collapse in the orthodox Dirac-von Neumann account.

[2] Bohm and Hiley 1993; Maroney and Hiley 1999.

wave packet that was initially ascribed to A divides into two parts and the position of A must fall, with equal probability, within either of those 'branches'. The measurement has the effect of BH-activating the branch picked off by the outcome. Although a subsequent spin measurement of B will also divide the wave packet ascribed to B into two parts, the only branch that will be BH-activated is that which was 'selected' by the previous spin measurement performed on A. Since C is, on the other hand, linked to A and B neither through its position nor through its spin, no measurement performed on C alone can possibly affect the behaviour of A or B.

Maroney and Hiley's account of quantum teleportation requires an appropriate measurement device (D) to be included in the representation. The device is initially ascribed a wave packet $\eta(x_0)$ where x_0 characterises the relevant 'pointer' position. When extended to include the apparatus, the initial representation (8.1.4) becomes

$$\frac{1}{2}\left(|\Phi^+>_{CA}|u_3>_B+|\Phi^->|u_2>-|\Psi^+>|u_1>-|\Psi^->|u_0>\right)|\eta(x_0)> \qquad (8.2.8)$$

A 'Bell interaction' is introduced, which generates a quantum potential that couples A, B, C with the device. More precisely, the Bell interaction effects a coupling of the x_0 coordinate of the apparatus to a 'Bell state' of C+A, leading to the final representation

$$\frac{1}{2}\left(|\Phi^+>_{CA}|u_3>_B|\eta_3(x_0)>+|\Phi^->|u_2>|\eta_2(x_0)>-|\Psi^+>|u_1>|\eta_1(x_0)>\right.$$
$$\left.-|\Psi^->|u_0>|\eta_0(x_0)>\right) \qquad (8.2.9)$$

where $|\eta_i(x_0)>$, i = 0,1,2,3 denote four non-overlapping 'position states' of the apparatus, each of which corresponds to one possible outcome of the Bell measurement. That measurement entangles all four systems in such a way that their subsequent behaviour is determined by their inherited "common pool of information". Whereas the initial form of the quantum potential was a product of two factors, one pertaining to C and the other to the A+B pair, the quantum potential produced by the interaction is a product of a C+A- and a B-related factor, both of which are functions of x_0. This transformation corresponds to the decoupling of A and B, whereby the dependence of the C factor on the coefficients a,b of the 'unknown state' X_C gets 'transferred' to the B factor of the final quantum potential. The Bell interaction results in the position variable x_0 of the device entering one of the four non-overlapping wave packets $\eta_i(x_0)$, thereby affecting the spin of B through the action of the quantum torque. That influence of the Bell measurement on B is non-local and instantaneous. The crucial point is that the branch in (8.2.9) that gets, as it were, highlighted by the apparatus coordinate remains BH-active, whereas all

three other branches become BH-passive. Bringing about that change of three branches from BH-active to BH-passive would be the essential purpose of the interaction within the Bell analyser. Of all bases that may be used for expanding the C+A+B+D initial state, the Bell basis would be special (only) in that it guarantees that, whilst only one of the branches remains active, the coefficients of the 'teleported' X_C state end up identical in every branch with those of the 'spin state' of B, up to a definite unitary **P**-group transformation.

Thus, quantum teleportation would boil down to BH-activity being transferred from one particle to another, through the mediation of a non-locally operating quantum potential. The eventual 'presence' in the B term of the "information carried by[1]" the a,b, coefficients would signal that transfer of BH-activity from Claire's particle to Bob's. However, which of the four branches remains BH-active after the completion of the Bell measurement cannot be guessed by Bob. Fortunately, the 2 bits' worth of information Alice provides suffice for him to know which of the four **P**-related spin states of B corresponds to the BH-active branch, perform the appropriate transformation and 'recover' the hitherto unknown former spin state of C.

Maroney and Hiley dismiss alternative accounts, based on so-called 'quantum information', on the grounds that the applicability of that concept is restricted to mixed cases. Their argument is that, if its von Neumann 'entropy' is the appropriate measure of the information that is potentially conveyed by a given quantum state, then since that entropy is zero in a pure case like $|X>$, regardless of the values of a and b, no such 'quantum information' is relevant to individual systems like A, B, C in a one-shot implementation of the quantum teleportation protocol. What is physically transferred from C to B should, in the authors' view, have "nothing to do with any kind of quantum information, which is, in any case, only 'information for us'[2]". Quantum teleportation should rather be thought of in terms of a transfer of "objective active information[3]", which they appear to conceive of as some kind of disposition or *potentia*, encoded in the wave function or state vector and "carried", as it were, by he a,b coefficients.

Maroney and Hiley assume that quantum teleportation cannot be understood without making an explicit reference to the interaction of the systems involved with an appropriate Bell-measuring device, 'position states' of which have to be included in the representation. However, whilst it is clearly necessary for A and C to interact in some suitable way in order to highlight any one 'branch' in (8.1.4) or (8.2.9), it is by no means clear why the basic Hilbert space representation *should* make an explicit reference to 'states' of the device involved. Position is also granted, in the Bohmian fashion, a suspicious

[1] Maroney and Hiley 1999.
[2] *Ibid.*
[3] *Ibid.*

privilege as an (the) indicator of the alleged BH-activity. At any rate, BH-activity (or lack thereof), the common pool idea and attendant aspects of the Bohm-Hiley view are all highly speculative, and the idea of a BH-activity transfer turns out to be just as immune to empirical disproof as the Jozsa-Penrose backwards-in-time transfer of their homespun 'quantum information'. What, here as elsewhere, appears to really count for any claim that quantum teleportation has been successfully achieved is Alice's 2-bit message. Whatever else may be added is, some might say, superfluous metaphysics that may or not, depending on one's taste and outlook, be regarded as providing further insight into the overall 'effect'.

8.3. From classical cryptography to quantum teleportation

Quantum teleportation as discussed in section 8.1 comprises four stages:

(i) Alice is provided with a qubit (**C**), the preparation of which remains unknown to her.

(ii) She jointly performs on **C** and **A** a special type of measurement. Alternatively, she first lets the qubits interact in a way that amounts, logically, to an XOR (**CNOT**), then measures them separately.

(iii) She informs Bob of the outcome of her measurement.

(iv) Bob operates on **B**, that must be suitably correlated to **A**, conditional on the information supplied in (iii), after what he can claim that his qubit is correctly 'described' using the ket, whatever it was, Claire had initially assigned to **C**.

Both the purpose: recovery of a 'message', and the means of achieving it: Alice and Bob's sharing a 'key' (in the form of pairs correlated in a definite way); a discriminating operation performed on **C+A**; the transmission of information that is necessary for retrieval of the message and Bob's eventually 'decoding' it bear a striking resemblance to the aims and methods of classical cryptography. Indeed, the four stages listed above are strongly reminiscent of a simple procedure for communicating secret messages[1], invented by the American engineer Gilbert Vernam in 1917.

[1] Stinson 1995.

Suppose Alice's task is to send a message to Bob whilst ensuring nobody else can possibly read it. For convenience, Alice's text is written in binary code (each letter, punctuation mark etc. being represented, say, by a 5-bit block: 00001 for 'A', 00010 for 'B' etc.) and the message forms a bit string M_A. Bob and Alice also possess their own copy of another and no shorter binary string K_{AB}, which they intend to use as a key for encryption and decryption. Alice takes the two strings M_A and K_{AB} and she performs XOR between each M_A bit and its match in the K_{AB} sequence: if the bits are different, the output bit is 1, and 0 otherwise ($0\oplus1=1\oplus0=1, 0\oplus0=1\oplus1=0$). This generates a string C_A that forms the encrypted message. Alice then either sends the string to Bob or she broadcasts it bit by bit. After he gets holds of C_A, Bob takes his own copy of K_{AB}, computes $C_A\oplus K_{AB}$ and recovers the intended message M_A. The exchange is secure insofar as an eavesdropper who does not know the key cannot read the message[1].

Despite similarities to (ii)-(iv), Vernam cryptography might seem to have little to do with quantum teleportation: after all, what is transferred is not a 'quantum state' but a finite sequence of bits that is *known* to Alice; the scheme involves no (pairs of) particles (not to say anything about entanglement); Alice and Bob perform no measurement. The parallel with teleportation can, however, be made more compelling if the original procedure is replaced with a stochastic variant. The relevant 'message' is no longer any given finite sequence of bits; rather, it is a probability distribution of bit values 0 and 1. These values may correspond to the outcomes of coin tosses, or to those of measuring a two-valued observable, e.g. σ_z. Given the initial preparation (a coin, a collection of identically prepared qubits...), repeated trials or measurements, yielding N random outcomes, will generate one of 2^N bit sequences. As $N \rightarrow \infty$, those sequences converge[2] toward a probability distribution, and it is that distribution Alice wants to convey to Bob. We shall suppose that the distribution is realised in Claire's preparation of an ensemble $\mathbf{E}_C$ of qubits: this preparation is such that measuring σ_z yields the value +1 for $100|a|^2$ percent of the qubits, and -1 for $100|b|^2$ percent of them, hence $p(+1/\mathbf{E}_C) = p_0 = |a|^2$ and $p(-1/\mathbf{E}_C) = p_1 = |b|^2$. The statistics for the ensemble can be calculated using the density operator

$$\rho_C{}^i = |a|^2 |0><0| + |b|^2 |1><1| \qquad (8.3.1)$$

[1] Vernam's cipher is also known as 'one time pad' because it is secure only if any one key K_{AB} is used for encryption and decryption of one message only.

[2] With all the due qualifications regarding such 'convergence'; see e.g. Howson and Urbach 1993 ; Howson 2000.

Alice holds a number of such **C** particles prepared in accordance with (8.3.1). Alice and Bob also hold **A** and **B** members of pairs of spin-½ particles, prepared in such a way that their z components of spin are always found opposite. More precisely, we shall assume that the ensemble of pairs is a mixture associated with the operator

$$\rho_{AB} = \frac{1}{2}|01><01| + \frac{1}{2}|10><10| , \qquad (8.3.2)$$

and $\rho_{CAB}{}^{i} = \rho_{C}{}^{i} \otimes \rho_{AB}$. (8.3.3)

Alice performs on randomly chosen **C** and **A** particles a 's/d measurement'. This measurement amounts to discriminating between **C** and **A** bits being the same ('s', coded as 0) or different ('d', coded as 1). The discriminating operation is the analogue of the XOR performed in the Vernam procedure. A suitable device can achieve it without Alice having to know anything about the bits involved. These results form a bit string Alice can then broadcast. If the value of the bit he receives is 0, Bob then performs a NOT operation (σ_x) on the matching **B** particle, whereas he does nothing if it is 1. In one case, the updated statistical operator is

$$\rho_{B,d} = tr_{CA}[(B_{CA}^{(d)} \otimes I_B)\rho_{CAB}{}^{i}(B_{CA}^{(d)} \otimes I_B)^{\dagger}] = \rho^{i}_{B/C} \qquad (8.3.4)$$

and in the other

$$\rho_{B,s} = tr_{CA}[(B_{CA}^{(s)} \otimes I_B)\rho_{CAB}{}^{i}(B_{CA}^{(s)} \otimes I_B)^{\dagger}] = \sigma_x \rho^{i}_{B/C}\sigma_x{}^{\dagger} \qquad (8.3.5)$$

where $B_{CA}^{(d)} = |01><01| + |10><10|$, $B_{CA}^{(s)} = |00><00| + |11><11|$ and $\rho^{i}_{B/C}$ is (8.3.1) acting upon $\mathbf{H}_B$.

Provided the experiment is repeated a large enough number of times, Bob can estimate the probability distribution of σ_z outcomes for the ensemble $\mathbf{E}_C$. This is clearly only made possible by the correlation between the bit values of **A** and **B**.

In such a probability distribution transfer (PDT), each of Alice's measurements supplies 1 bit that must be communicated to Bob. This is just half of what is required for quantum teleportation. Implicit in choosing (8.3.1) is the selection of a context for experimental trials: the density operator (8.3.1) corresponds to a definite probability distribution with respect to just one observable (σ_z), and the operations performed in PDT refer to bit values associated with eigenkets of that observable. Besides, the expansion of (8.3.2) comprises none of the off-diagonal terms that are typical of entanglement and signal 'EPR correlations'. Thus, the effectiveness of PDT

hinges on **A+B** pairs forming an ordinary 50%-50% mixture (8.3.2), in contrast to quantum teleportation where it hinges on maximal entanglement (Bell kets). This suggests writing a density operator for **A+B** pairs as

$$\rho_{AB,\lambda} = \frac{1}{2}[\,|01><01| + |10><10| - \lambda(|01><10| + |10><01|)\,]\,. \qquad (8.3.6)$$

PDT is realised by taking (8.3.1) as the input and letting $\lambda = 0$. The ordinary mixture $\lambda = 0$ gives rise to no 'quantum correlation'. At the other extreme, the singlet ($\lambda = 1$), like the other 'Bell states', exhibits the highest degree of such correlation. For any given value of the parameter $\lambda \in [0,1]$ and an input $|X>_C$ the output density operator ρ_B^f, after Bob has performed the appropriate operation, is

$$\rho_B^f = |a|^2|0><0| + |b|^2|1><1| + \lambda\,(ab^*|0><1| + a^*b|1><0|). \qquad (8.3.7)$$

The input is (statistically) retrieved only if $\lambda = 1$: entanglement is required for teleportation to be successfully achieved. If $\lambda = 0$ the output is a 'classical' mixture, whose predictive potential is clearly not equivalent to that of a ket $|X>$. There is no 'teleportation' since the off-diagonal contribution to the input statistical operator does not appear in the output.

Let us briefly examine how the probability of retrieval of the input by Bob is affected by the value of λ. The most general form of $|X>_C$ is

$$|X>_C = \cos(\frac{\theta}{2})|0> + e^{i\phi}\sin(\frac{\theta}{2})|1>\,,$$

where the range of values of θ and ϕ is continuous. Using (8.3.7), the probability of interest is

$$p(\theta,\lambda) = tr_B(\rho_B^f|X><X|) = \frac{1}{4}\,[\lambda(1-\cos 2\theta) + \cos 2\theta + 3] \qquad (8.3.8)$$

This probability does not depend on the relative phase ϕ (it is indifferent to whether the amplitudes are real or complex). As an affine function of λ, $p(\theta,\lambda)$ has a maximum $p(\theta,1) = 1$. For $\lambda \neq 1$ the probability, as a function of the angle θ, has a minimum $p(\frac{\pi}{2},\lambda) = \frac{1}{2}(\lambda+1)$, and the lowest value $p(\frac{\pi}{2},0) = \frac{1}{2}$ obtains for PDT and a 'maximally quantal' input i.e. for a superposition of basis kets $|0>$ and $|1>$ with equal probabilities. This shows that PDT is fundamentally inadequate for transferring the predictive potential embodied in a Hilbert space vector (note that, for all λ, the probability is unity for the orthogonal inputs $|0>$ ($\theta=0$) and $|1>$ ($\theta=\pi$)).

Just as the quantity of information that can be experimentally acquired 'on' a single qubit equals 1 bit, the amount that can be acquired in measuring a qubit pair is just 2 bits. Alice's s/d measurements supply only 1 bit, which is enough for Bob to be able to recover (8.3.1), but only half of what a binary preparation yields to experiment. Pairs of qubits can be prepared in such a way that measuring a given observable separately on each member yields a given outcome with probability 1 (such preparations are denoted by ket products). Alternatively, a binary preparation may be completely specified by reference to the relative outcomes of measuring two incompatible observables. Suppose that a pure case binary preparation is such that the 'experimental question' $Q_{opp}^{(z)}$: "Are the values of the z spin component found opposite when that component is measured separately on **A** and **B**?" will have a 'yes' answer for every **A+B** pair from a given ensemble. The corresponding $\mathbf{H}_A \otimes \mathbf{H}_B$ ket must then be a linear combination $\alpha|01> + \beta|10>$, where α and β are two complex numbers related by the condition $|\alpha|^2 + |\beta|^2 = 1$. A 'yes' answer to $Q_{opp}^{(z)}$ yields just 1 bit; it is insufficient to determine α and β. Requiring that the question $Q_{opp}^{(x)}$: "Are the values of the x spin component of **A** and **B** found opposite?" be *also* answered 'yes' adds one more bit, making it possible to specify (up to a meaningless phase factor) the $\mathbf{H}_A \otimes \mathbf{H}_B$ vector associated with the preparation: consistency between the linear combination written above and $\lambda|0'1'> + \mu|1'0'>$ demands that $\beta = \lambda = -\mu = -\alpha$, i.e. that the preparation be ascribed the ket $|\Psi^->$. The other three Bell kets similarly obtain by requiring a 'yes' answer to a $Q^{(x)}$ *and* a $Q^{(z)}$ question: $|\Phi^+>$ by requiring that the values of both spin components be found the same for **A** and **B**; and $|\Psi^+>$ ($|\Phi^->$) that the values of the x (z) component be the same, those of the z (x) component opposite[1]. The four Bell kets exhaust the range of possibilities[2] when it comes to sorting ensembles of qubit pairs on the basis of the s/d dichotomy[3]. This sheds some light on the reason why Bob eventually has to

[1] A similar remark is made by Zeilinger 1997,1999. This author also claims, however, that all there is to qubits is that they are 'carriers of information', and also that it is necessary that the 'information content' of a system [?] scale with its size. Zeilinger's Bohrian inclinations notwithstanding, this suggestion comes dangerously close to ontologising information. It is perhaps worth mentioning that statements to the effect that physical systems possess or carry information abound in the literature; for example: "In the absence of connections between the detectors and the source, a particle *has no information about* how the switch of its detector will be set until it arrives there...It would seem essential for each particle *to be carrying instructions* for how its detector should flash for either of two possible switch settings it might find upon arrival." (Mermin 1990b).

[2] They form an orthonormal basis of 2-qubit Hilbert space.

[3] Aspects of the same vs. opposite distinction in QIT are discussed by Maroney 2000. Unsurprisingly, sameness and oppositeness measurements in the quantum setting differ from their classical counterparts essentially because of non-orthogonality in the former. Non-orthogonality implies that even if there is a fact of the matter as to any two preparations

perform one of four **P**-transformations: those transformations exhaust the possible compositions (products) of σ_x and σ_z, which are the two reference observables for the above questions. Nor is it a coincidence that in the CNOT-based account, Alice's measurement involves (whether explicitly or not) those same two incompatible observables.

Generalising PDT to input kets rather than classical mixtures calls for the resources of maximal entanglement, i.e. ρ_{AB} must be of the form (8.3.6) with $\lambda = 1$. The quantum teleportation of a qubit appears to be such a generalisation. Any information transferred would then boil down to the 2 bits Alice must acquire and then communicate to Bob, in contrast to the single bit PDT requires. In either case, Bob has to wait for Alice's message, conveyed via ordinary channels, to reach him wherever he is – patience at a distance...This conclusion is a far cry from a vision of individual 'states' being disassembled to be later reconstructed at a remote location. Many, indeed, will resist it on the grounds that quantum teleportation, as it is usually conceived, is a 'one-shot' operation: it is believed to 'work' with a single instance of Claire's qubit and a single **A+B** pair. The issue actually boils down to that of the meaning of the various state assignments at stages (i) to (iv) for each of the agents involved. Claire's assignment of $|X>_C$ reflects her maximal knowledge of her preparation. On the other hand, Bob and Alice, who have no prior information about that preparation, are justified in assigning the density operator $\rho_C{}^i = \frac{1}{2}\mathbf{I}_C$. No measurement Alice could have performed on **C** alone at any time between t_0 and t_1 (see diagram, Section 8.2) could possibly have refuted that assignment – for all Alice and Bob can know, Claire might just as well have supplied Alice with a member of some maximally entangled pair. After t_2, their epistemic situation with respect to **C** has not changed: they would then still assign to that qubit a density operator $\rho_C{}^f = \frac{1}{2}\mathbf{I}_C$, consistently with $tr_{AB}(|\Xi><\Xi|)$. On the other hand, because all of them know that **A** and **B** are members of a maximally entangled pair Claire, Alice and Bob assign them the statistical operators $\rho_A{}^i = \frac{1}{2}\mathbf{I}_A$ and $\rho_B{}^i = \frac{1}{2}\mathbf{I}_B$. After *she* learns that Bob performed on his qubit the **P**-transformation that corresponds

---

counting as 'same' or 'opposite' in the appropriate sense, performing measurements associated with different basis choices cannot give us any certainty as to whether samples from either preparation will both 'pass' or fail' ('same' case) or only one will ('opposite'). A clear-cut same vs. opposite discrimination requires the relevant logical gates to operate in a properly specified basis, making any such discrimination a contextual matter. One consequence of this is that quantum algorithms cannot in general simply be broken down into separate and locally operating sub-algorithms.

to Alice's two-bit message, Claire can confidently update her **B**-related assignment to the same ket $|X>_B$ she had initially assigned 'to' **C**. Alice and Bob, on the other hand, can only trust the formalism of quantum theory to correctly account for a 'transfer' to which they contribute as blind agents. Claire can later on persuade Alice and Bob that 'teleportation' was actually achieved by asking Bob to perform a 'yes-no measurement' associated with the operators $\mathbf{O}_X = |X><X|$ ('yes') and $\mathbf{I} - \mathbf{O}_X$ ('no'). Then, provided Alice did supply Bob with the appropriate 2 bits and he operated on **B** accordingly, that measurement should yield a 'yes' answer. But Claire only would know *which* yes-no measurement to choose. If Bob instead picked off *that* measurement by a sheer guess, his obtaining a 'yes' answer would not warrant his claim that **B** is at that time 'in the state' $|X>$: had he measured the same observable on **B** before t_1, he would have had a 50% chance of obtaining a 'yes' answer, since $tr_B(\rho_B{}^i\, \mathbf{O}_X{}^{(B)}) = \frac{1}{2}$.

The mystifying weirdness of quantum teleportation resolves into a straightforward account provided Hilbert space vectors or density operators are regarded as predictors and properly referred to the epistemic status of the parties involved at the various stages in the procedure. The objectivity of the 'transfer' then boils down to the mutual consistency of the assignments made by all of those parties.

8.4. Dense coding, entanglement swapping

Let, again, a binary preparation correspond to the ket $|\Psi^->_{AB}$. One particle (**A**) of each prepared pair goes to Alice, the other one (**B**) to Bob. Acting upon **A** in a suitable way, Alice can send to Bob a 2 bits' worth of information, hence twice as much as might be expected given the mere 1 bit that a spin measurement performed on **A** can yield. The two-bit message may, for instance, be one of the four DNA bases A, C, G, T. Alice and Bob agreed to make each of the four letters correspond one-to-one to a **P**-operation performed on **A**: σ_z for the letter A, σ_x for C, $\sigma_x\sigma_z$ for G and the identity $\mathbf{I}_A$ for T. To encode 'C', Alice takes particle $\mathbf{A}_k$ from the pair No.k and subjects it to a NOT gate (σ_x): the resulting $(\mathbf{A}+\mathbf{B})_k$ ket is $|\Phi^->_{AB}$. Once this is done, $\mathbf{A}_k$ is returned to its box and that box sent to Bob. After he receives it, Bob Bell-measures $\mathbf{A}_k+\mathbf{B}_k$ and the resulting ket, as intended by Alice, is $|\Phi^->_{AB}$. This is enough for him to learn that Alice's message is 'C'. This works because,

given the choice of a Bell ket, there is a one-to-one correspondence between Alice's operation, hence the letter she encoded, and the result of Bob's Bell measurement: if the initial 'state' is $|\Psi^-\rangle_{AB}$, Bob is sure to obtain (a result that corresponds to) $|\Psi^+\rangle$ for A, $|\Psi^-\rangle$ for T, $|\Phi^+\rangle$ for G and $|\Phi^-\rangle$ for C.

Such *(super)dense coding*[1] hinges on the initial preparation being associated with a Bell ket, for Alice's operation on **A** consists in a 1-1 mapping of one such ket into another one from the same set ($I_A \otimes I_B$, $\sigma_x^{(A)} \otimes I_B$, $\sigma_z^{(A)} \otimes I_B$ and $\sigma_x\sigma_z^{(A)} \otimes I_B$ form a group that maps the Bell basis onto itself). Each operation affects the Hilbert space vector that is associated with the relevant pair in the following way: identity aside, Alice's operation alters the initial preparation by turning one or both of the 'same' questions that define the associated Bell vector[2] into an 'opposite' question, and the other way round. If the preparation is a singlet, $\sigma_x^{(A)}$ leaves the initial $Q_{opp}^{(x)}$ unchanged but enforces a 'Yes' answer to the alternative question $Q_{same}^{(z)}$. Similarly, $\sigma_z^{(A)}$ changes the initial $Q_{opp}^{(x)}$ into $Q_{same}^{(x)}$ whilst leaving the initial $Q_{opp}^{(z)}$ unchanged, and $\sigma_x\sigma_z^{(A)}$ turns $Q_{opp}^{(x)}$ and $Q_{opp}^{(z)}$ into $Q_{same}^{(x)}$ and $Q_{same}^{(z)}$ respectively. Notice that σ_x (σ_z) effects a change in the question that corresponds to the 'complementary' σ_z (σ_x) observable: 2 bits are what it takes to specify any Bell ket; 2 bits can also be acquired (dense coding) by identifying a **P**-transform of a Bell ket.

An ontological reading would have particle **B** be affected in an obscure and seemingly rather spooky manner[3] by an operation performed on the remote particle **A**. This overlooks the fact that no definite (pure) state can be meaningfully ascribed to **B** or **A** prior to Alice's operation (any possible outcome of measuring a given two-valued observable on **A** or **B** separately is equally likely; mixtures associated with the reduced density operators on $\mathbf{H}_A$ and $\mathbf{H}_B$ are improper). Given the above-mentioned mapping, Alice's intervention changes nothing to this. Her performing any **P**-group transformation alters the *s/d relation* between the possible outcomes of measurements that may be carried out on **A** and **B** separately. Because a 'Yes' answer to any of the $Q^{(x)}$ and $Q^{(z)}$ questions provides no clue regarding the value that may be found if some observable is actually measured on **B**, the statistics of measurements to be performed on **B** alone are not altered in any way by Alice's intervention. Whether any single particle **B** undergoes any change as a result of Alice's operation is a question that cannot be given an operational meaning – no more indeed than enquiring about *the* state

---

[1] Bennett and Wiesner 1992.

[2] See Section 8.3.

[3] An account of dense coding along the lines of Jozsa's interpretation of quantum teleportation raises the same questions and prompts essentially the same remarks as in Section 8.2.

of the **B** particle, as long as the preparation is properly 'described' using a Bell vector.

As a final illustration of the necessity to be mindful of the predictive and statistical significance of 'state' vectors, let us briefly consider one last 'effect' involving pairs associated with a Bell ket. After their production, Alice picks out any two pairs **A+B** and **C+D**. She labels them identically e.g., $(\mathbf{A+B})_{16}$ and $(\mathbf{C+D})_{16}$, keeps one member of each, say $\mathbf{A}_{16}$ and $\mathbf{C}_{16}$, and sends the other two ($\mathbf{B}_{16}$ and $\mathbf{D}_{16}$) to Bob. All of the particles remain isolated until Alice or Bob decide to operate on them. Since any two $(\mathbf{A+B})_k$ and $(\mathbf{C+D})_k$ pairs are mutually uncorrelated, all predictions regarding part or all of the qubits $\mathbf{A}_k$, $\mathbf{B}_k$, $\mathbf{C}_k$ and $\mathbf{D}_k$ can be derived from the single product ket

$$|\Omega_{ABCD}\rangle = |\Psi^-_{AB}\rangle\otimes|\Psi^-_{CD}\rangle$$
$$= 2^{-1/2}(|01\rangle_{AB} - |10\rangle_{AB})\otimes 2^{-1/2}(|01\rangle_{CD} - |10\rangle_{CD}). \qquad (8.4.1)$$

Let us expand that ket, with the bit slots in the alternative order ACBD, as

$$|\Omega_{ACBD}\rangle = \frac{1}{2}(|0011\rangle - |0110\rangle - |1001\rangle + |1100\rangle) \qquad (8.4.2)$$

and write it finally

$$|\Omega_{ACBD}\rangle = \frac{1}{2}(|\Phi^+\rangle_{AC}\otimes|\Phi^+\rangle_{BD} - |\Phi^-\rangle\otimes|\Phi^-\rangle$$
$$- |\Psi^+\rangle\otimes|\Psi^+\rangle + |\Psi^-\rangle\otimes|\Psi^-\rangle). \qquad (8.4.3)$$

At any later time she likes, Alice takes stored **A** and **C** particles that bear the same label, e.g. $\mathbf{A}_{16}$ and $\mathbf{C}_{16}$, and lets them interact inside a Bell device. The output corresponds, with equal probability, to any one of the Bell kets $|\Phi^\pm\rangle_{AC}$, $|\Psi^\pm\rangle_{AC}$. Given (8.4.3), Alice is led to the *same* assignment with respect to Bob's matching $(\mathbf{B+D})_{16}$ pair. Alternatively, she may perform $\mathbf{CNOT}_{AC}$, followed by a Hadamard transform on **A**, so that $|\Omega_{ACBD}\rangle$ evolves into

$$|\Gamma_{ACBD}\rangle = \tfrac{1}{2}(|00\rangle_{AC}\otimes|\Phi^+\rangle_{BD} - |01\rangle\otimes|\Psi^+\rangle - |10\rangle\otimes|\Phi^-\rangle + |11\rangle\otimes|\Psi^-\rangle), \qquad (8.4.4)$$

and then measure σ_z on $\mathbf{A}_m$ and $\mathbf{C}_m$ separately. There again, each ordered pair of outcomes uniquely corresponds to a Bell ket for the remote $(\mathbf{B+D})_m$.

Since the location where $\mathbf{B}_m$ and $\mathbf{D}_m$ are held by Bob is arbitrarily distant from Alice's, the effect has been interpreted as the creation at a distance of

previously nonexistent correlations. The above procedure, known as *entanglement swapping*[1], would "entangle two quantum systems that have never interacted directly with each other[2]". However, the remarks of Section 8.1 also apply. Bob *has* to be informed by Alice of *which* **A+C** pairs she Bell-measured and *which* result she got if he is to select those **B** and **D** particles that correspond to a given Bell ket, say $|\Phi^+>_{AC}$, perform e.g. appropriate spin measurements on **B**'s and **D**'s from that subset and obtain statistics that are consistent with the assignment of a given Bell ket. Failing to receive that information, Bob would be at a loss to corroborate in any way the claim that some of his **B** 'end up entangled' with the similarly labelled **D**. If he were to perform spin measurements on randomly chosen $\mathbf{B}_k$ and $\mathbf{D}_k$ particles, the statistics he would obtain would not reveal any correlation. Thus, no evidence of **B+D** entanglement can be exhibited unless subsets of $(\mathbf{B+D})_k$ particles are selected by Bob on the basis of Alice's information, and the correlations are computed for one subset thus selected. The telltale statistics are characteristic of *selected subsets* of Bob's records, that must be consistent with Alice's records and (8.4.3). In fact, Bob could just as well have performed measurements on **B**'s and **D**'s *before* Alice did anything with her qubits, and discarded his particles immediately afterwards[3]. After Alice had made her measurements, he would then still have been able, using the information she provided, to gather as much evidence of maximal entanglement as if he had just kept his particles in store and patiently waited for her message to carry out any operation on **B**'s and **D**'s bearing the same label.

---

[1] Zukowski *et al.* 1993 ; Bose *et al.* 1998 ; see Pan *et al.* 1998 for an experimental realisation with polarised photons.
[2] Bose *et al.* 1999.
[3] Peres 1999.

9

Further developments

9.1 Gleason's theorem and positive operator-valued measures

A linear predictive scheme hinges on two basic assumptions:

(i) Simple experimental questions correspond to orthonormal bases over some appropriate metric vector space or, equivalently, to complete sets of orthogonal one-dimensional projectors over some Hilbert space.

(ii) The probability of a given outcome must not depend on any particular choice of embedding set for the corresponding projection operator.

Probability assignments associated with sets of linearly independent but not necessarily orthogonal vectors would be mere expressions of indifference, thereby nullifying the predictive value of the mathematical scheme. Why this is so can be seen from the crude following argument. If, to each of n linearly independent but non-orthogonal $\mathbb{R}^n$ vectors v_i, there corresponds a unique probability p_i, then rotating one v_k of those vectors whilst holding the others fixed should leave the probability p_k unchanged, as long as v_k does not end up collinear to any other v_i. Nearly all the unit sphere vectors (except for those that fall on a ray spanned by any one the n-1 remaining v_i) will therefore correspond to probability p_k. Since this must be the case for any k, it follows that all of the p_i will be the same i.e. $p_i = \frac{1}{n}$.

Assumption (ii) makes perfectly good sense within a scheme in which the predictors are fundamentally conventional. Gleason's theorem[1] is a rigorous expression of the consistency of probability valuation with the basic structure of a linear predictive scheme: if predictors associated with definite yes-no answers to experimental questions are constrained, in accordance with (i) and (ii), by the projective structure of Hilbert space, some suitable statistical operator acting upon $\mathbf{H}_d$ exists such that the probabilities of measurement outcomes take the usual trace form (no structure of interest happens to be pinned down if the Hilbert space is two-dimensional). More precisely, if we

[1] Gleason 1957; Hughes 1989 (appendix).

postulate the existence of a function p from the set of one-dimensional projectors into the real interval [0,1] that satisfies $\sum_{i=1}^{d} p(\mathbf{P}_i) = 1$ ($d \geq 3$), then for any such function p there exists a statistical operator ρ such that, for all *i*, $p(\mathbf{P}_i) = tr(\rho \mathbf{P}_i)$. No further assumption is required for the result to hold e.g. there is no need for p to be continuous or for $\mathbf{H}_d$ to be over the complex field. Paul Busch recently proved[1] a new version of Gleason's theorem as part of an attempt to meet common objections to von Neumann's 'no go' argument against non-contextual hidden variables[2]. The upshot of reformulating the original argument in terms of 'unsharp' observables (POVM) rather than self-adjoint operators is a strengthening of von Neumann's result: there are no dispersion-free valuations on the set of 'effects'. Positive operator-valued measures (POVM) were introduced in the 1980s with a view to generalising a projective account of measurement[3]. In the standard approach, possible measurement outcomes are identified with the eigenvalues of self-adjoint operators (**P5**) or complete sets of eigenkets thereof. Given a suitable statistical operator ρ, consistent probability assignments require that eigenprojectors $\mathbf{P}_i$ corresponding to the eigenvalues a_i of some observable **A** form a resolution of the identity operator **I**: $\sum_i \mathbf{P}_i = \mathbf{I}$ (so that the probabilities $p_\rho(a_i) = tr(\rho \mathbf{P}_i)$ add up to 1). In the new approach, it is assumed instead that sets $\{\mathbf{E}_\mu\}$ of positive-semidefinite operators ($< u | \mathbf{E}_\mu | u > \geq 0$ for all $|u> \in \mathbf{H}$) that sum up to identity ($\sum_\mu \mathbf{E}_\mu = \mathbf{I}$) qualify[4] as ('unsharp') observables. The probabilities of μ-labelled outcomes are computed using a POVM version of **PR**: $p_\mu = tr(\rho \mathbf{E}_\mu)$. There is no a priori constraint on the number of indices and the $\mathbf{E}_\mu$'s need not be mutually orthogonal. Whereas projection-valued measures are representations of ideally 'sharp' observables, the set of positive operators (technically known as *effects*[5]) is intended to correspond to a more spread out, hence 'more realistic' range of experimental questions.

In Busch's version of Gleason's theorem, a frame function p from positive operators $\{E_\mu\}$ to [0,1] is postulated, such that $p(E_\mu) \geq 0$ and $\sum_\mu p(E_\mu) = 1$. The theorem says that there must exist a density operator ρ such that $p(E_\mu) = tr(\rho E_\mu)$. The assumptions behind the proof are straightforward

---

[1] 1999, published in Busch 2003; see Caves *et al.* 2004 for an independent, equivalent proof.
[2] Von Neumann 1955. See Mermin 1993 for a review.
[3] Kraus 1983; Peres 1993; Busch *et al.* 1995.
[4] If $|u>$ is a normalised ket, it is necessary and sufficient for $< u | E_i | u >$ to represent a probability that E_i be an operator bounded between the null and the unit (identity) operators.
[5] Kraus 1983.

POVM analogues of (i) and (ii). One of the most striking features of Busch's proof is its simplicity, which contrasts with the rather formidable technical resources mobilised by Gleason's. No less remarkable is the fact that Busch's version encompasses the two-dimensional case, to which Gleason's fails to apply. Qubits are no longer 'special'; it just happens that all outcome statistics for ideally sharp measurements on two-level systems can be mimicked using the resources of some non-contextual hidden variable theory. Finally, the greater simplicity and wider scope of the Gleason-Busch theorem suggests POVM might contribute in an invaluable manner to illuminating the foundations of quantum theory.

9.2 'Quantum' Bayesian update, complex numbers and ⊗-composition

On the standard account, the 'final' eigenket, after application of the reduction rule **P7**, may be almost orthogonal to the initial ket, as if the actual outcome was barely related to the 'initial conditions'. In contrast, the ordinary probability calculus gives a very simple account of the update of one's knowledge after a piece of evidence e_k has been gathered. The probability distribution p(h) for some given hypothesis h is expanded as a sum

$$p_{initial}(h) = \sum_i p(h \wedge e_i) = \sum_i p(h/e_i)p(e_i), \qquad (9.2.1)$$

where $\{e_i\}$ is a relevant data set and $p(e_i)$ the prior probability assigned to the data e_i. The term that corresponds to the acquired data is just picked off the sum, so that the updated probability of h given the acquired information is

$$p_{final}(h) = p(h/e_k) = p(h \wedge e_k)/p(e_k) \qquad (9.2.2)$$

It has been suggested that reduction might be "an artifact of a problematic [Hermitian operator-based] representation[1]" and that using POVMs might help to make the connection tighter between classical 'Bayesian' update and the change of statistical operator prompted by the acquisition of fresh information. Given an initial density operator ρ, the one that is associated with the final situation after the measurement[2] of a POVM $\{E_\mu\}$ is not unique but only

[1] Fuchs 2001.

[2] The measurement of a POVM on a system can be represented as a projective measurement on an auxiliary system ('ancilla') that formerly interacted with it. Information is then gathered indirectly by measuring the ancilla rather than the system of interest. This possibility is a simple consequence of Hilbert space tensor product composition and of the **PR** rule in its trace form, which together imply that whatever pertains to the auxiliary system can be 'traced out' (Holevo 1973a; Davies 1978).

required to be of the general form $\rho_\mu = \frac{1}{tr(\rho E_\mu)} \sum_i A_{\mu i} \rho A_{\mu i}^\dagger$, where $\sum_i A_{\mu i}^\dagger A_{\mu i} = E_\mu$ (there is no constraint on the number of indices μ, nor do the A operators have to be self-adjoint). The reduction rule (P7') of the standard formalism is a special case[1] of the $\rho \rightarrow \rho_d$ transition (following Fuchs[2], the subscript *d* – for 'data' – will from now on be used in place of μ).

Trying to work out as close a 'quantum' analogue as possible to the classical expansion (9.2.1), one faces the difficulty that the initial ρ cannot simply be expanded in the form $\sum_d p(d)\rho_d$, where $p(d) = tr(\rho E_d)$. However, rewriting ρ as $\rho^{1/2} I \rho^{1/2}$ with $I = \sum_d E_d$ leads to $\rho = \sum_d p(d) \tilde{\rho}_d$, where the operators ρ_d and $\tilde{\rho}_d$ happen to have the same eigenvalues. The $\rho \rightarrow \rho_d$ transition can then be regarded as comprising two stages: first, a $\rho \rightarrow \tilde{\rho}_d$ transition, followed by a formal readjustment (unitary transformation) that is required to complete the change from ρ to ρ_d. A projective measurement amounts to a mere readjustment[3] whereas one's performing a measurement on one member of an 'entangled pair' prompts an update of one's beliefs regarding both members, without any further readjustment[4]. Thus a POVM treatment leads to thinking of the effect of acquiring information through measurement as a variant of classical Bayesian update satisfying constraints imposed by a non-commutative framework. Writing $\rho = \sum_d p(d) \tilde{\rho}_d$, then $p(h) = \sum_d p(d) p(h/d)$, where $p(h/d) = tr(\tilde{\rho}_d E_d)$, and the analogy with (9.2.1) is complete. Some light is also thrown on the assumption (ii) of the previous Section: it is the Hilbert space analogue of the simple remark that, while updating from $p_{initial}(h)$ to $p_{final}(h) = p(h/d)$ depends on p(d), the transition itself does not depend on how the remainder of the outcome set to which the data *d* belongs has actually been partitioned.

Holding a view of Hilbert space vectors or density operators as predictors rather than descriptions of objective states of affairs confronts us with the possible meaning of *unknown* states. This is not an idle question, for references to 'unknown quantum states' pervade the QIT literature. The case of quantum teleportation is easily settled: as argued in detail in chapter 8, it requires one party at least (Claire) to hold maximal information about the **C**-preparation i.e. to make at the outset a pure case assignment in the form of the Hilbert space vector $|X>_C$. However, other mention of 'unknown states'

[1] Fuchs 2001, Nielsen and Chuang 2000.
[2] Fuchs 2002.
[3] Fuchs 2001,2002, part 6.
[4] *Ibid.*

may not so easily be disposed of. Proponents of a Bayesian interpretation of the probability calculus have encountered a rather similar problem: whereas the frequentist or the propensitist have no difficulty thinking of unknown probabilities, the expression is an oxymoron for the Bayesian, whose job starts with an assignment of non-zero priors. Fortunately, Bruno de Finetti's representation theorem[1] disposes of the need to refer to any problematic notion of unknown probability, replacing it with the assignment of an exchangeable distribution in multi-trial space[2]. Caves *et al.*[3] have adapted de Finetti's approach to the quantum setting. Under the assumption that frequency data only are relevant to updating a density operator assignment, these authors show that the updating rule is similar to that for classical Bayesian update[4]. As long as two experimenters are not inflexible in their beliefs, the accumulation of new data may force them to revise their assignment until they reach closer agreement. The account dispenses with the need to think in terms of objective, initially unknown 'quantum states' that would reveal more and more of their identity as evidence accumulates.

An intriguing aspect of this 'quantal' version of the de Finetti theorem is that it fails for *real* Hilbert spaces[5]. Whether or why complex numbers are actually necessary for the SAQM to work at all remains a vexed issue. Would the probabilistic machinery of quantum theory retain anything of its effectiveness if Hilbert space were chosen real rather than complex? Peremptory claims notwithstanding – e.g. "complementarity...*demands* that the probability amplitudes of quantum mechanics be complex[6]" – attempts to answer the question have been rather inconclusive. Thus, Caticha[7] just confesses his inability to justify the need for amplitudes to be complex rather than real. William Wootters[8] showed that the way points are distributed over the probability simplex is uniform if the associated amplitudes are complex, whereas the points turn out to be concentrated toward the border of the simplex if the amplitudes are chosen real (and toward the centre for quaternions). However, it remains to be proven that quantum amplitudes *must* be complex if uniformity over the Hilbert hypersphere is to translate into uniformity (maximal randomness) over the probability simplex[9].

---

[1] de Finetti 1964.

[2] Multi-trial probabilities are said to be *exchangeable* if they are invariant under any permutation of their arguments; see Kyburg and Smokler 1964, Howson and Urbach 1993, Howson 2000.

[3] Caves *et al.* 2001; Fuchs 2002.

[4] Schack *et al.* 2001; Fuchs 2002.

[5] See Fuchs 2001,2002 for a counterexample.

[6] Wheeler 1990, p.12 (italics added).

[7] Caticha 1998.

[8] Wootters 1980.

[9] Wootters (1980) also tried to derive **PR** using a variational method. Extremising the gain of Shannon information as an **A**-measurement yields an outcome a_i, a squared-modulus form

More recently, an intriguing relationship has been unveiled between the question of 'complexity' and another basic issue: that of tensor product composition. Let $\{E_\mu\}$ and $\{F_\nu\}$ be two POVM measured separately on systems labelled 1 and 2 respectively, the relevant Hilbert spaces $\mathbf{H}_1$ and $\mathbf{H}_2$ being d_1 and d_2-dimensional. $\{E^\nu_\mu\}$ is also introduced[1] as the POVM measured on 1 conditional on having previously obtained the outcome ν when measuring $\{F_\nu\}$ on 2, and the other way round for $\{F^\mu_\nu\}$ (then, $\sum_\mu E^\nu_\mu = \mathbf{I}$ for all ν and $\sum_\nu F^\mu_\nu = \mathbf{I}$ for all μ). The crucial assumption is now made[2] that the joint probability $p_{\mu\nu}$ associated with either ordered pair $S_{\mu\nu}$ of operators (E_μ, F^μ_ν) or (E^ν_μ, F_ν) should not depend on which set of ordered pairs $S_{\mu\nu}$ is embedded in (the assumption is clearly reminiscent of the 'non-contextuality' assumption (ii) behind the Gleason-Busch theorem). Fuchs's theorem[3] says that, given the existence of a function f from the set of such ordered pairs into the interval [0,1], satisfying the condition $\sum_{\mu\nu} f(S_{\mu\nu}) = 1$, there must also exist a linear operator σ on the tensor product space $\mathbf{H}_1 \otimes \mathbf{H}_2$, such that for any ordered pair of locally measurable POVM (E,F) : $f(E,F) = \mathrm{tr}(\sigma(E \otimes F))$. Moreover, σ happens to be unique if both Hilbert spaces are defined over the field of complex numbers, because the set $\{E_\mu \otimes F_\nu\}$ then forms a complete basis for Hermitian operators on $\mathbf{H}_1 \otimes \mathbf{H}_2$. Whether uniqueness holds only in the complex case remains open. This is plausible though since, when the field is complex, the operator space of the tensor product is isomorphic to the tensor product of the original operator spaces. The isomorphism does *not* hold, however, in the real field case, because the dimensionality $\frac{1}{2}d_i(d_i+1)$ of the space of symmetric operators[4] on a real Hilbert space $\mathbf{H}_i$ is strictly less than the dimensionality d_i^2 of the vector space of Hermitian operators over the base space in the complex case, and hence insufficient for σ to be uniquely specified. Fuchs's assumptions are not strong enough to establish that σ must be a density operator[5]. Nevertheless, it is significant that one should get so close to the correct trace form of probability valuation in the binary case from a purely local account of measurement, which closely parallels Wootters's 'local accessibility' thesis: "any set of measurements which are just sufficient for

---

for the expression of the probability $p_\Psi(a_i)$ is recovered. However, the derivation, made under the assumption of *real* Hilbert space, fails in the complex case.

[1] Fuchs 2002.

[2] *Ibid.*

[3] Fuchs 2002.

[4] Symmetric operators are the real-case analogues of Hermitian operators.

[5] See Fuchs 2002, equ. (61) for a counterexample.

determining the states of the subsystems are, when performed jointly, also just sufficient for determining the state of the combined system[1]."

9.3 A simple axiom set for quantum theory

An important contribution to our understanding of quantum theory has been made by Lucien Hardy[2], who shows that the SAQM can be derived from five basic axioms. The axioms in question, which make no prior reference to Hilbert space or set a priori limits on the mutual compatibility of physical quantities, give rise to a POVM formulation of quantum theory. **P1-P7** follow if the operators are restricted to being Hermitian. One of the most remarkable aspects of this reconstruction is that both quantum theory qua linear predictive scheme and the classical probability calculus satisfy four of the axioms. In fact, dropping a continuity requirement that is crucial to the fifth axiom is all it takes to switch from the SAQM to classical probability theory. Hardy's approach makes a strong case for claiming that, consistently with the view advocated in the whole of Part B of the present dissertation, "[q]uantum theory, when stripped of all its incidental structure, is simply a new type of probability theory[3]".

To motivate his approach, Hardy introduces a typical set-up allowing measurements to be performed on suitably prepared samples (the axioms by themselves make no reference to any such set-up). This includes (i) a preparing device that releases samples when it is in the 'on' mode (the device comprises a knob for selecting the preparation); (ii) a transforming device such that any released sample undergoes a systematic transformation as it passes through it (a knob setting on the device specifies the applied transformation); and (iii) a measurement apparatus, which has a knob for selecting the measurement being performed. As far as the statement of the basic axioms and their consequences are concerned, the role of (ii) can be reduced to the identity transformation. For a given setting of the knob and provided the preparing device is on, the measurement yields any one of a fixed number of non-null outcomes (one, and only one, of a number of properly installed detectors must then fire). The outcome of a measurement is null (no detector fires) if the preparation device is 'off'. The relative frequency of a given output (including the null output) can be calculated from a series of runs of the experiment with fixed settings of all three devices. The purpose of the first axiom (HA1) is precisely to connect probabilities of measurement outcomes to

[1] Wootters 1990, p. 44.
[2] Hardy 2001a.
[3] Hardy 2001b, p.1.

the actual gathering of relevant data. This is done in the usual way by identifying those probabilities with limiting relative frequencies: the relative number n_k of instances of a given outcome a_k in N runs of a measurement for a given setting of all the devices (and provided the release button of the preparation device is pressed) is assumed to have a definite limit as $N \to \infty$, which limit yields the probability of that particular outcome.

A given setting of the preparation device determines the probability of each non-null outcome for any measurement performed on a released sample. Hardy chooses to designate by the word *state* whatever, in the preparation effected by (i), is relevant to predicting any such probability. Unfortunately, saying that we "ascribe states to preparations[1]" or, worse still, that the device prepares "systems in some state[2]" is a very misleading way of expressing our capacity to work out all the predictive consequences of a given choice of knob setting for (i). A 'state', as introduced above, is *not* meant to provide e.g. a list of attributes of each individual (sample) system released by the preparation device. The primary – indeed, the sole – purpose of the formalism to be derived from the axioms is predictive rather than descriptive.

It is reasonable to assume that there exists a minimum number[3] K of probabilities that suffice for specifying the state i.e. the 'predictive content' of the preparation. To each of those probabilities will correspond a definite measurement, and the state will be exhausted by listing the corresponding probabilities p_i, i=1 to K. A major step toward the representative framework to be derived is Hardy's choice, reminiscent of Destouches's[4], to write those probabilities as a K-dimensional column *vector*:

$$\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ . \\ . \\ . \\ p_K \end{pmatrix}.$$

Besides, it is assumed that sets of 'states' exist, such that those states can be distinguished from each other within any such set by a 'single shot' measurement. If each state in the set gives rise to a different outcome for a given setting of (iii), then performing that measurement will suffice to identify the state, hence the preparation (of course, this requires that whoever is provided with a sample thus prepared is informed of which measurement he should perform). Let N then be maximum number of states in any such set, i.e.

---

[1] Hardy 2001b, p.2.

[2] *Ibid.*, p.3.

[3] Hardy refers to K as the "number of degrees of freedom". We shall not use this terminology, for it suggests a parallel with classical mechanics that is almost incongruous in the current context.

[4] Section 5.3.

adding any more state would imply some ambiguity in identifying members of the extended set. In quantum theory, the number $N \neq K$ will be shown at a later stage to coincide with the dimension of the predictively relevant Hilbert space.

The remaining four axioms (HA2 to HA5) impose structural constraints on the predictive scheme to be derived. We have no reason a priori to expect K and N to be equal, although we do expect and shall assume that there exists a structural relationship between those two numbers. According to Hardy's second axiom (HA2), such a relationship exists regardless of the types of physical systems involved. Hardy himself ascribes that link to "a certain constancy in nature[1]", but one can alternatively motivate it by requiring consistency between the conditions set to the distinguishability of observable outcomes within a vector space structure (the N side) and those set to the specification of an appropriate preparation (the K side). Hardy includes in (HA2) a clause to the effect that K, as a function of N, takes the smallest value that is consistent with the axioms. If the mode of preparation constrains observations in such a way that some outcomes, though possible a priori, are never actually observed, then (HA3) the maximum number of distinguishable states i.e. the dimension of the relevant space can be reduced to a lower value $M < N$ (the probability of occurrence of a given outcome within a set should not depend on whether that set is embedded in some larger set).

Before going any further let us, for the sake of illustration, think of the preparing device as a machine that releases either red or green 'balls'. Any given preparation i.e. knob setting of the device is fully specified by giving the two probabilities p_R and p_G of observing a red or a green ball respectively. Accordingly, K = 2 and a 'state' vector $\vec{p} = \begin{pmatrix} p_R \\ p_G \end{pmatrix}$ can be introduced ($p_R + p_G \leq 1$ in general if balls are not systematically released by the device). Since the maximum number of observational outcomes that may be distinguished in this experiment is N = 2 (any released ball is either green or red), K = N in this particular example. The two vectors $\vec{p}_1 = \vec{p}_R = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\vec{p}_2 = \vec{p}_G = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ will correspond to a setting of the preparation device such that the balls it releases are respectively all red ($\vec{p}_1$) or all green ($\vec{p}_2$). These two 'states' are special in so far as, unlike any $\vec{p}$ whose components are strictly less than 1, they cannot be simulated by probabilistic mixtures of other states. Hardy calls *pure* such states as denoted by $\vec{p}_1$ and $\vec{p}_2$. If the preparing device is 'off', no ball is

[1] Hardy 2001a, 6.14.

released and $p_R = p_G = 0$, hence the 'null state' corresponding to $\vec{p}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. $\vec{p}_0, \vec{p}_1$ and $\vec{p}_2$ are extremal in the K-dimensional space of acceptable states (the set of allowed states is the 'simplex' defined by the condition $p_R + p_G \leq 1$). A measurement performed on any released ball consists in finding out which colour it is. If we ascribe a vector $\vec{r}_1 = \vec{r}_R = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to a basic measurement of the 'red observable', the probability that such a measurement will yield a 'yes' answer for a preparation specified by the vector $\vec{p}$ is $p_R = \vec{r}_1 . \vec{p}$. Similarly, a 'green' measurement corresponding to $\vec{r}_2 = \vec{r}_G = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ will have a positive outcome with probability $p_G = \vec{r}_2 . \vec{p}$.

This extremely basic formalism is readily generalised to higher dimensions. Since any probability that can be ascertained (HA1) by measurement is determined by the setting of the preparation device, that probability must be a function f of $\vec{p}$ (that function will generally differ from one measurement to another). Suppose that Bob selects at random with probability *w* a preparation associated with the vector $\vec{p}_A$ or, with probability[1] *1 – w*, another setting of the preparing device that corresponds to the vector $\vec{p}_B$. Unless she is informed of the selection procedure, Alice will not identify either preparation through her own measurements. Rather, she will ascribe the results she obtains to a fictitious preparation associated with a vector $\vec{p}_C$, whilst only Bob can write explicitly the probability function as the convex combination $f(\vec{p}_C) = wf(\vec{p}_A) + (1 - w)f(\vec{p}_B)$. For each of the K measurements used to determine $\vec{p}$, the relation $p_{C,i} = wp_{A,i} + (1 - w)p_{B,i}$, i = 1 to K, holds between the probabilities. It must therefore hold for the 'state vectors' themselves: $\vec{p}_C = w\vec{p}_A + (1 - w)\vec{p}_B$, suggesting that the function f is linear. Hardy shows that it is and that the probabilities of measurement outcomes are given by the dot (scalar) product $\vec{r} . \vec{p}$ of the appropriate $\vec{r}$ and $\vec{p}$ vectors[2].

Since the set S of allowed states is convex and bounded (the components of $\vec{p}$ are obviously confined within the interval [0,1]), it admits an extremal set: this is the set of those vectors, like $\vec{p}_1$ and $\vec{p}_2$ above, that cannot be written as

[1] This probability must here clearly be understood as a measure of degree of belief, throwing doubt on the necessity of Hardy's commitment to a frequency interpretation (HA1); see Schack 2002.

[2] Hardy 2001a, Appendix 1.

a convex sum of other vectors in S (the null state is excluded from the set of 'pure' states). Hardy takes it as a "driving intuition" in his approach that "pure states represent definite (non-probabilistic) *states of the system*[1]". Why should this be the case, however, since 'states' were introduced in the first place as a conventional listing of probabilities based (HA1) on repeated runs of test experiments? In accordance with the predictive aim of the proposed formalism, Hardy's 'pure states' should rather be thought of as providing 'reference preparations' in a somewhat similar sense as the 'eigenpreparations' of Section 6.1. That Hardy's states can be indifferently[2] represented using $\vec{p}$-vectors or $\vec{r}$-vectors is another motive for resisting the temptation to ascribe states to individual systems and to uncritically reify them. The fourth axiom (HA4) is concerned with circumstances when "a preparation device ejects its system in such a way that it [the system] can be regarded as made up of two subsystems[3]" i.e. with binary preparations. If unary preparations involving 1-systems only are characterised by a number N_1 of distinguishable states, and unary preparations involving 2-systems only by a number N_2, the number of distinguishable states associated with 1+2 preparations cannot be less than N_1N_2. According to (HA4), for all binary preparations N is *equal* to N_1N_2. Probabilities for measurements performed on (i)-samples (i = 1 or 2) can be listed as the $K_i = K(N_i)$ components of a vector $\vec{p}^{(i)}$. Sets of pure states $\{\vec{p}_m^{(1)}\}$, with m = 1 to K_1, and $\{\vec{p}_n^{(2)}\}$, with n = 1 to K_2 must then exist, from which K_1K_2 linearly independent matrices $p_{mn}^{(1+2)}$ can be constructed. To each $\vec{p}_k^{(i)}$ there corresponds a measurement $\vec{r}_k^{(i)}$ that picks it off, and the probability p_{mn} that measuring $\vec{r}_m^{(1)}$ and $\vec{r}_n^{(2)}$ separately will yield a 'yes' outcome for both is the product of $p_m^{(1)}$ and $p_n^{(2)}$. The p_{mn} form the K_1K_2 entries of a matrix $p^{(1+2)}$ and the corresponding 'joint state' can be expressed as a convex combination of K_1K_2 linearly independent 'product states' associated with the $p_{mn}^{(1+2)}$ matrices. These considerations suggest (second part of HA4) that the number K of real parameters that are necessary and sufficient for defining a state in the binary case is also just equal to the product K_1K_2. Hardy further motivates equality by asserting, in accordance with Wootters' local accessibility idea[4], that "[t]here should not be more entanglement [and, should we add, no less] than necessary[5]" to carry out our predictions successfully.

---

[1] *Ibid.* (italics added).
[2] *Ibid.*,2.
[3] *Ibid.* 6.16.
[4] Wootters 1990.
[5] Hardy 2001b, p.10.

For a fixed choice of basis set in the N-dimensional vector space of distinguishable states, there are a number $\frac{N!}{m!(N-m)!}$ of m-dimensional subspaces and a number u_m of measurements that are necessary and sufficient to completely characterise a 'state' confined to that subspace. The number K can be expanded as

$$K = N u_1 + \frac{N(N-1)}{2} u_2 + \ldots$$

From (HA3-HA4), $K(N^2) = K^2(N)$ and $K(N+1) > K(N)$, hence the only polynomial in N that satisfies those conditions is $K(N) = N^r$, with r an integer.

In quantum theory, if the relevant Hilbert space is of dimension N, then $N^2 - 1$ independent real numbers are required to specify a density matrix. If the predictors associated with the two experimentally accessible 'sides' of a binary preparation are M and N-dimensional respectively, then the number of independent probabilities that are necessary and sufficient to determine the density matrix pertaining to the whole is

$$(M^2 - 1) + (N^2 - 1) + (M^2 - 1)(N^2 - 1) = (MN)^2 - 1.$$

This is an expression of Wootters' 'local accessibility' thesis[1]: quantum theory as a predictive scheme makes an optimal use of the information supplied by measurements performed separately on subsystems, including the correlations one may extract from a careful comparison of outcome lists. If g(N) is the number of parameters needed to specify a normalised 'state' (in Hardy's sense) when the relevant preparation is such that the maximum number of distinguishable outcomes of an appropriate reference measurement is N, the optimality condition can be written

$$g(M) + g(N) + g(M)g(N) = g(MN).$$

This condition is satisfied both by complex Hilbert space quantum theory, where $g(N) = N^2 - 1$, and by classical probability theory, where $g(N) = N - 1$. On that basis, Wootters (1990) conjectured that any theory that satisfies the optimality condition might have to be such that $g(N) = N^r - 1$, with r a non-negative integer. But this is just the expression that follows from Hardy's axioms: $K(N) = g(N) + 1 = N^r$. Adhering to the simplicity clause of (HA2) leaves us with only two kinds of predictive scheme: those for which r = 1 or 2. It is at least conceivable that the SAQM itself might be embeddable in some $r > 2$ framework, just as the classical r = 1 framework is shown to be embeddable in 'quantum theory' (r = 2). The restriction is sensible, however, if it is acknowledged that quantum theory functions as a r = 2 predictive scheme, by

[1] Wootters 1990. The product MN is of course just the dimension of the tensor product space $\mathbf{H}_1 \otimes \mathbf{H}_2$.

virtue of the mathematical structure implied in the selection of square integrable solutions of the Schrödinger equation. On the other hand, there is as yet no basic framework for theoretical physics that would appear to call for the resources of a r > 2 scheme. We are thus left with two cases. K = N i.e. r = 1 corresponds to a linear representation of the classical probability calculus, in which the number of parameters needed to characterise a preparation is equal to the maximum number of distinguishable 'states'. In the ball example, specifying the state boils down to supplying the probabilities p_R and p_G (K=2), while the number of distinguishable pure 'basis states' is obviously N=2: at most two colours can be reliably distinguished in a single observation of any given ball. That K and N must be equal just means that no observation can be performed that cannot be reduced to identifying the colour of the ball. On the other hand, there is a priori no reason why, for all conceivable types of system and preparation, the maximum number of observable outcomes for a given setting of the measuring device should be equal to the number of probabilities required to completely specify the preparation (or more accurately to determine all predictions for a given setting of the knob of the preparing device). The complex Hilbert space structure of the SAQM is known to satisfy $K = N^2$ i.e. r = 2. *Real* Hilbert spaces are actually ruled out[1] by (HA4). It is a structural property of those spaces that $u_1 = u_2 = 1$, $u_{i\geq 3} = 0$ i.e. $K = N(N+1)/2$, hence $K > K_1K_2$: the specification of states in the binary case requires more than can provide information gathered separately on subsystems[2] and working out correlations between the two lists of outcomes, thus belying local accessibility. In other words, complex numbers appear to be required for setting up an optimal linear predictive scheme.

Let us briefly outline how the SAQM can be recast so as to contrast it with the ball example. All predictions stemming from a 'qubit' preparation can be calculated from a statistical operator or its realisation $\rho = \begin{pmatrix} a & b \\ b^* & d \end{pmatrix}$ in a given basis, e.g. that of the eigenkets of σ_z. Since, then, $a = p_{z,+} = \text{prob}(\text{'}\sigma_z\text{=+1'})$, $d = 1 - a$ and the real and imaginary parts of b are equal to $p_{x,+} - ½$ and $½ - p_{y,+}$ respectively, knowing the probabilities of '+' outcomes of a spin measurement in the three directions x ,y and z is sufficient to completely determine the preparation. If normalisation is not imposed, one further probability $p_{z,-}$ is required, and it is completely equivalent to represent the state

[1] A counterexample is given by Wootters 1990, p.44.

[2] Hilbert spaces over the quaternions, for which $u_1 = 1$, $u_2 = 4$ ($u_2 = 2$ in the complex case), $u_{i\geq 3} = 0$ would require strictly *less*: $K < K_1K_2$.

by ρ or using the 4-component vector $\vec{p} = \begin{pmatrix} p_{z,+} \\ p_{z,-} \\ p_{x,+} \\ p_{y,+} \end{pmatrix}$. In the normalised case, the allowed states can be represented as a ball sitting inside the unit cube in the coordinate system ($p_{x,+}$, $p_{y,+}$, $p_{z,+}$). The surface of that ball is an exact analogue of the Bloch sphere representation of qubit 'states' (the points on the surface of the ball represent extremal (pure) states). If normalisation is not assumed, all the predictively useful information encoded in a N-dimensional density matrix ρ will be equivalent to that supplied by a vector whose components are $K = N^2$ probabilities. This equivalence can be traced back to the algebraic fact that any Hermitian operator that can be represented by a N×N matrix can also be written as a linear combination (with real coefficients) of $K = N^2$ projection operators $\mathbf{P}_1,\ldots,\mathbf{P}_K$. Writing these K operators as the components of a vector $\vec{P}$, Hardy's 'state' vector can be expressed as $\vec{p} = \mathrm{tr}(\rho\vec{P})$. The most general expression of probability in quantum theory is $\mathrm{tr}(\rho\mathbf{A})$, where $\mathbf{A}$ is a positive operator associated with the measurement. A measurement vector $\vec{r}$ is introduced[1] through the relation $\mathbf{A} = \vec{r}.\vec{P}$, leading to the same expression of probability as in the red-or-green ball case: $p = \vec{r}.\vec{p}$. The fundamental difference between the 'classical' and the 'quantal' cases resides in the composition of the sets of allowed states i.e. of $\vec{p}$ and $\vec{r}$ vectors. The distinctive composition of those sets is shown to be deducible from all of the axioms, without any extra assumption.

One last axiom (HA5) is introduced in order to select the r = 2 case. Quantum theory obtains if it is additionally required that, for any dimension N, a *continuous* and reversible transformation[2] exists along a (continuous) path connecting any two pure states (the intermediate states are also pure). The set of classical pure Hardy states (K = N) is discrete (vertices of a triangle in the ball example), whereas e.g. qubit pure states form a continuous set (points on the surface of the Bloch sphere). The continuity requirement is, in Hardy's approach, the key difference[3] between 'quantum' and classical probability theory: it is simply not possible in the latter for a pure state to be continuously transformed into another, and (HA2) implies that the only alternative is quantum theory ($K = N^2$). It is somewhat ironical that *classical* probability theory should be characterised by a necessity to 'jump' between

---

[1] Hardy 2001a, 5, 8.7.

[2] A continuous transformation is understood as built up from transformations, each of which differs only infinitesimally from the identity.

[3] "It is rather striking that the difference between classical probability theory and quantum theory is just one word." Hardy 2001b, p.10.

pure states, whereas the basic trait that singles out quantum theory among the schemes that satisfy Hardy's other axioms (HA1 to HA4) would be the existence of *continuous* transformations! Following his "driving intuition", Hardy suggests that the necessity to 'jump' from a classical state corresponding to, say, a ball *being* in one box to that corresponding to its *being* in another would just (?) reflect our crude partitioning of possibilities into what just *appears* to us to be clear-cut alternatives. The range of (objective) possibilities might stretch infinitely beyond such crude alternatives, as befits a world that is supposedly suffused with that famous 'quantum uncertainty'...How such speculations relate at all to the kind of basic framework Hardy sets up is hard to see. They sound, rather, like a desperate attempt to secure a minimum of 'physical content' to a framework that is conspicuously independent of assumptions relating to features of the physical world, or to systems as identifiable parts thereof. When it comes to assessing the basic differences implied by the acceptance or rejection of the fifth axiom, the relevant comparison is between two distinct but closely related types of probabilistic schemes (sharing four axioms out of five) and not between e.g. quantum and classical variants of mechanics. The radical view of quantum theory that emerges from Hardy's work is one "in which states (pure or mixed) are the analogues of classical probability distributions, and are not seen as a property of the individual system, but [relate to] a specific way of preparing the systems[1]", or more accurately to all the predictive (and only predictive) consequences of any such preparation.

For any choice of r, the effect on $\vec{p}$ of the transformation device can be described by a K×K real matrix. In the r = 2 case, this is equivalent to a linear transformation of the statistical operator ρ. The most general transformations that are consistent with all the axioms are linear and completely positive maps[2] on the space of operators on Hilbert space, such that the time evolution does not increase the trace (or equivalently the normalisation coefficient in the pure case). Unitary evolution obtains if it is required that the corresponding transformation should also be invertible and that it should preserve the trace, whereas 'reduction' means net trace decrease signalled by the increase of the von Neumann entropy. None of these two special cases, however, corresponds to any necessary constraint on the generalised formalism. How the representation should be updated when new information is acquired is subject to the same basic constraints in classical (probability) and quantum theory, but the updating rules reflect the structural features of each particular

[1] Werner 2001.

[2] Since the mid-1990s such maps (Busch *et al.* 1995, Nielsen and Chuang 2000) have become the stock in trade of the QIT theorist. Typical concerns include looking for the optimal map that satisfies some information-processing criterion, or showing that no such map is consistent with certain information-processing requirements.

scheme. Thus, the POVM alternative to reduction outlined in Section 9.2 can be regarded as a r = 2 variant of classical (r =1) Bayesian update.

9.4 Can information-theoretical concepts and methods illuminate the foundations of quantum theory?

The technological benefits of developing a 'quantum mathematical theory of communication' were recognised only during the last decade of the twentieth century[1]. QIT underwent fast development in the 1990s following the publication of two influential papers: that of Bennett *et al.* (1993) on quantum teleportation and Shor's (1994) on 'quantum factoring'. The extremely ambitious goal of some QIT theorists is "to arrive at a set of principles like energy and momentum conservation, but which apply to information, and from which much [and indeed, according to the most radical, the whole] of quantum mechanics could be derived[2]." Expressions of that goal and of major insights of QIT, e.g. that "according to the laws of quantum mechanics, Nature remarkably is able to process information exponentially more efficiently than can be achieved by any classical means[3]" usually conflate references to an objective physical world ('Nature') with capacities and demands – e.g. of computational efficiency – that hardly made any good sense irrespective of cognitive agents, their interests and purposes (e.g. to carry out and finalise a computation).

John A. Wheeler's speculations have been a source of inspiration for a number of prominent QIT theorists, like Deutsch, Fuchs or Wootters. Wheeler tirelessly advocated[4] his conviction that the bedrock of theoretical physics will be found to consist in simple truths about information (*It from bit*). His concern with information ties in with his participatory view of quantum theory, which gives the observer a key role in ensuring the self-consistency of his bootstrapping picture of reality as emerging through 'quantum networking[5].' It is much more unclear what a basic informational grounding for theoretical physics could amount to in the context of an entirely observer or IGUS[6]-free picture of reality. Nonetheless, there is a widely shared feeling inside the QIT community that the new field is actually heading toward the fulfilment of part at least of Wheeler's grand vision. As David Deutsch puts it,

---

[1] The emergence of quantum computing can be traced back to some thought-provoking papers by Feynman 1982,1986.
[2] Steane 1998, p.42.
[3] Ekert and Jozsa 1998 (italics added).
[4] Wheeler 1982,1983,1988,1990.
[5] Wheeler 1988.
[6] 'Information Gathering and Utilizing System' (Gell-Mann and Hartle 1990).

> Of John Wheeler's 'Really Big Questions', the one on which most progress has been made is *It From Bit?* – does information play a significant role at the foundations of physics? It is perhaps less ambitious than some of the other Questions, such as *How Come Existence?*, because it does not necessarily require a metaphysical answer. And unlike, say, *Why the Quantum?*, it does not require the discovery of new laws of nature : there was room for hope that it might be answered through a better understanding of the laws as we currently know them, particularly those of quantum physics. And this is what has happened: the better understanding is the quantum theory of information and computation[1].

Pace Deutsch there is very little, if anything, in what has been achieved in QIT so far that lends any clear support to Wheeler's *It From Bit*. QIT has, from its inception, been chiefly a matter of adapting traditional IT engineering problems e.g., that of optimising data transmission over a suitable channel or that of enhancing algorithmic efficiency, to a setting in which vectors and matrices, subject to specific constraints, are used in place of ordinary probability distributions[2]. Sweeping statements about 'quantum information' and its virtues notwithstanding, the endeavour, as it stands, appears to offer little prospect of illuminating the aim and structure of an essential part of theoretical physics. Besides, there is something circular in attempting to ground quantum physics on a novel treatment, never mind a new kind of information, if the latter owes its 'non-classical' features from its being precisely abstracted from the basic formalism of quantum theory.

---

[1] Deutsch continues: "How might our conception of the quantum physical world have been different if *It From Bit* had been a motivation from the outset? No one knows how to derive *it* (the nature of the physical world) from *bit* (the idea that information plays a significant role at the foundations of physics), and I shall argue that this will never be possible. But we can do the next best thing: we can start from the qubit." (Deutsch 2003).

[2] For example, the 'quantum' noiseless coding theorem (Schumacher 1995) is a rather straightforward adaptation of Shannon's. Let X be a set of symbols x_k (k=1 to M) whose probability of occurrence is $p(x_k)$. A N-symbol message, where N is very large, will typically contain $Np(x_i)$ occurrences of x_i. The number of typical sequences is $\frac{N!}{(Np(x_1))!...(Np(x_M))!} \approx 2^{NH(X)}$, where $H(X) = -\sum_{x_i} p(x_i)\log p(x_i)$ is the Shannon function. H(X) is said to represent the quantity of information per symbol or, by abuse of language, the 'information content' (in bits) of the source. As N tends to infinity, atypical sequences occur with negligible probability and there remains only $2^{NH(X)}$ equiprobable typical sequences. Every such sequence can be labelled with a number and that number be sent in place of the N-letter message : the original message is compressed from NlogM bits to NH(X) bits. This compression is optimal (Shannon's theorem). In Schumacher's version, the source is assumed to 'prepare' systems as specified by the statistical operator ρ. Although an assembly of N qubits can be generally ascribed a N-qubit subspace, the Hilbert space characterisation of the assembly can be projected onto a subspace of lower dimension $2^{NS(\rho)}$, where $S(\rho)$ is the von Neumann entropy. In close parallel with Shannon's theorem, in which H(X) is a measure in bits of the resources required to convey the statistical features of the source X (described by a given probability distribution), the von Neumann entropy S is a measure (expressed in terms of a number of qubits) of the resources required to convey the statistical features of a preparation specified using the operator ρ.

Christopher Fuchs, who is currently one of the most dedicated and consistent proponents of the 'strong QI' program, has been briskly defending[1] his conviction that one of the priorities of theoretical research in physics should be to look for a small set of information-theoretic principles from which the whole of quantum theory would be seen to emerge in an almost self-evident manner. His personal driving intuition is that the whole of quantum physics might follow from a single 'crisp' statement like 'gathering information generally implies a disturbance to the information that may be gained by some other party[2]'. If quantum theory is 'physical' as it is meant to be, it would be through that property, distinctive of a 'quantum world', whereby observers are kept from ever gathering more data than Hilbert space vectors allocated to a pure case preparation allow them to. Quantum theory would be nothing but the essentially unique framework for prediction that could accommodate that property. Fleshing out the idea of a fundamental trade-off between information gain and 'disturbance' requires being able to quantify such gain. The three information measures Fuchs considers[3] are mutually consistent in that larger 'mutual' Shannon information (1) corresponds to a smaller probability of error (2) and to smaller statistical overlap (3). On the other hand, these measures do not mesh to the point that any measurement that is optimal for any one of them is also optimal for the others. On the 'disturbance' side, the extent to which a prior assignment is affected by one's gain of information can be variously thought of in terms of a distance between the initial and the final epistemic situation (a disturbance measure will typically vanish if the initial and final assignments coincide). Drawing on the 'no-broadcasting[4]' result, Fuchs[5] suggests that non-commutativity (of density operators) might be the key feature behind the sort of basic trade-off between information gain and disturbance he is after. Indeed, a 'which is which' gain of information regarding the identity of two pure case preparations implies no disturbance just in case the two density operators are orthogonal, hence clonable. However, no such connection exists in the mixed case between broadcastability and 'no disturbance'. It remains to be seen whether such considerations can lead to some simple reciprocal relationship (as simple, say, as a Heisenberg-type inequality) between the two kinds of measure. Until this is achieved, Fuchs programme cannot be regarded as more than theoretically motivated wishful thinking. Besides, the endeavour falls short of addressing pressing questions.

---

[1] Fuchs 2001,2002.

[2] Such a statement is central to quantum cryptology, and in particular to the implementation of secure schemes of quantum key distribution; see Bennett and Brassard 1984; H.K. Lo in Lo *et al.* 1998; Nielsen and Chuang 2000.

[3] Fuchs 1996,1998. These are also, in a definite sense, measures of distinguishability; see end of Section 6.2.

[4] Barnum *et al.* 1996 ; see footnote 1 p.144.

[5] Fuchs 1998.

First and foremost, why did physicists have to start probing the world on microscopic scales and look for an appropriate framework for atomic physics to hit upon what would later be (properly?) recognised as an expression of trade-offs that are necessarily part and parcel of our attempts to gather and exchange information within a 'quantum world'? And why *quantum* for that matter? Why should the impressive empirical adequacy of quantum mechanics and its derivative frameworks (quantum field theories in particular) relate to any such trade-offs? And why should that kind of trade-off have anything to do with the finiteness of an *action* quantum?

It is a consequence of the structural impossibility to establish by a single empirical test the veracity of a statement about a 'quantum' preparation[1] that performing a measurement on one of the two arms of a binary preparation cannot be used to implement a scheme that would allow information to be conveyed superluminally. If one could, just by measuring an individual system, find out 'its' ket or density operator, then one could at the very least decide whether one is dealing with a pure or with a mixed case. Then, given 1+2 pairs prepared e.g. as singlets, it would be possible for Bob, by performing just one measurement on 2, to deduce whether or not Alice had carried out a measurement on the matching particle 1: the answer would be 'yes' if his measurement identified a pure case, and 'no' for a mixed case. Since this could be achieved regardless of the spatial separation between Alice and Bob, this possibility would provide them with a strategy for communicating 'faster than light'. Thus there is a sense in which, because of the above limitation, quantum theory precludes any form of instantaneous or superluminal transfer of information between two parties via entangled pairs, if any such transfer is meant to be achieved through measurements performed on either pair member[2].

Another implication of the standard formalism is the non-existence of a systematic procedure whereby, starting from two uncorrelated preparations with density operators σ and ρ, one would end up with a binary preparation generally exhibiting entanglement and such that the reduced (marginal) density operators for the two 'parts' would both be ρ. Such 'broadcasting', the restriction of which to pure cases ($\sigma\otimes\rho \rightarrow \rho\otimes\rho$) is 'cloning', is known to be ruled out in the quantum-theoretical setting, unless the preparations to be broadcast (cloned) happen to be associated with mutually commuting density matrices (orthogonal kets)[3]. 'No cloning' had, a decade prior to broadcasting,

---

[1]Various proofs have been given that any such attempt is doomed to failure. See e.g. Dieks 1988, Peres 1988, Jaeger and Shimony 1995, Busch 1996. All that one can afford is to increase the confidence level by testing more than one sample.

[2]For a critique of 'quantum no-signalling' proofs in historical perspective, see Kennedy 1995, and Lo *et al.* 1998; Nielsen and Chuang 2000 for typical QIT arguments.

[3] Barnum *et al.* 1996, Dieks 1982, Wootters and Zurek 1982. See note 1 p.144.

been pointed out as a rather trivial though intriguing consequence of the linearity[1] of the standard formalism of quantum theory. In fact, one may wonder whether much of the psychological impact of 'no cloning' did not arise because of the very choice of the word[2] *cloning*. Coupled with a view of Hilbert space vectors or density matrices as representatives of ontologically construed 'states', this results in a problematic picture in which all the objective physical attributes of a single individual system would be systematically reproduced in a different embodiment[3].

Finally, unconditionally secure 'bit commitment' is precluded by Hilbert space structure and ⊗-composition. In a bit commitment protocol one party, Alice, supplies an encoded bit to another, Bob. The encoding procedure is such that it should not allow Bob to ascertain the value of the bit until Alice provides him with additional information. This she does at the time she 'reveals' the bit value she committed initially. Can a bit commitment protocol be set up, such that Bob can confidently assume that the value Alice reveals to him actually is the one she committed earlier on? In 1984, Bennett and Brassard[4] showed that, by making cunning use of two-qubit entanglement, Alice can cheat without Bob being able to detect it. Lo and Chau[5], and independently Mayers[6] subsequently showed that a cheating strategy based on entanglement is always possible in principle.

Consider the following three 'no go': (i) performing a measurement on one system only cannot instantaneously or 'superluminally' have any influence on future predictions pertaining to any remote and independent system (modulo a suitable criterion of independence); (ii) no broadcasting; (iii) no unconditionally secure bit commitment. Clifton, Bub and Halvorson (CBH) have recently argued[7] that these three 'no go' are, if they are taken as postulates, sufficient to deduce a very general formulation of quantum theory, encompassing all variants of quantum field theory. Taken together, they are shown to entail the mutual commutation of observables pertaining to different 'parts', general non-commutativity for the observables that pertain to any one part considered on its own, and the pervasiveness of non-classical correlations associated with

---

[1] The original 'no cloning' argument of Wootters and Zurek (1982) makes no reference to unitarity.

[2] A. Peres suggests that another initial source of attraction towards 'no-cloning' is its loose connection to the question of 'superluminal influences' (identified as such in QIT related work despite the general absence of relativistic considerations *stricto sensu*), adding: "As it often happens in science, these things were well known to those who know things well." (Peres 2002)

[3] A similar remark could just as well be made about the expression 'quantum *teleportation*'.

[4] Bennett and Brassard 1984.

[5] Lo and Chau 1997.

[6] Mayers 1997.

[7] Clifton *et al.* 2003.

entanglement. As it turns out, those three features are all one needs for setting up an abstract quantum theory as the authors view it.

In accordance with a general trend in theoretical QIT work, CBH opt for a thoroughly C*-algebraic treatment, the generality of which encompasses all forms of quantum theory, from the most familiar to the most exotic. By a unital C*-algebra[1] is meant a Banach *-algebra over the field of complex numbers, such that involution (*) is related to norm by $\|A^*A\| = \|A\|^2$. A C*-algebra is spanned by positive operators that constitute its so-called 'effects'. If * is the adjoint operation and $\| \ \|$ the standard operator norm, then the basic Hilbert space structure of Galilean-relativistic quantum mechanics obtains[2]. A crucial step taken by the authors consists in defining a *state* of a C*-algebra as any positive and normalised linear functional on the algebra. In standard quantum theory, such a state obtains by selecting a positive and unit trace operator ρ and defining a linear and positive functional via $tr(\rho A)$ for any A in the bounded algebra of linear operators. The physical, or more accurately operational significance of such a state lies entirely in its supplying "a catalog of the expectation values of all observables[3]". The physical observables are to be represented by self-adjoint elements of the relevant unital subalgebras, where the latter encompass everything about the relevant systems that is amenable to prediction. To what extent such elements qualify as observables is to be justified on group-theoretical grounds, as outlined in Section 2.2, but the authors have no concern with that particular issue.

The C*-algebraic notion of state just introduced does not presuppose countable additivity, hence Gleason's theorem is not applicable. The absence of any such presupposition is primarily due to the fact that no assumption needs to be made a priori regarding the availability of an infinite and convergent sum of orthogonal projectors. As a matter of fact, there are pure (dispersion-free) states in the C*-algebraic sense that cannot be represented by Hilbert space vectors or density operators. On the other hand, it so happens that, for any C*-algebraic state, *some* representation can be set up in which that state admits a vector representative (whether in the pure or mixed case). Eventually, "every abstract C*-algebra has a concrete faithful representation as a norm-closed *-subalgebra of the bounded algebra of linear operators, for some appropriate Hilbert space. So there is a sense in which C*-algebras are no more general than algebras of operators on Hilbert spaces – apart from the fact that, when working with an abstract C*-algebra, one does

[1] See e.g. Kadison and Ringrose 1997.

[2] Every von Neumann algebra is a C*-algebra, but the converse does not hold.

[3] Clifton *et al.* 2003.

not privilege any particular concrete Hilbert space representation of the algebra[1]."

It is a somewhat neglected fact[2] that an operator formulation can be given for both classical and quantum mechanics. Classical observables being real-valued functions on a 6N-dimensional phase space, pure states then correspond to phase space points, warranting the tacit identification of the C*-algebraic predictive notion of state with the usual kinematical one[3]. Probability measures on phase space are introduced to tackle those cases in which detailed knowledge of kinematically relevant values is not available. As the authors point out, "not only does every classical phase space representation of a physical theory define a C*-algebra, but, conversely, behind every abstract abelian C*-algebra lurks in its function representation a good old-fashioned phase space theory[4]." We would therefore be justified in "treating a theory formulated in C*-algebraic language as classical just in case its algebra is abelian", and believing that "a necessary condition for thinking of a theory as a *quantum* theory is that its C*-algebra be non-abelian[5]." The condition is not sufficient, however, for non-commutativity by itself implies nothing as to the existence or not of such non-classical correlations as are associated in the standard formalism with entanglement.

A conveniently "larger" C*-algebra $A \vee B$ can be generated from two C*-algebras A and B and used as the appropriate arena for a C* state representation of joint (predictive) statements pertaining to both 'arms' of a binary preparation. A criterion for regarding those arms as mutually independent, for all predictive purposes, regardless of whether or not A-related observables commute with B-related ones is that, for any 'state' $\rho_1 \in A$ and $\rho_2 \in B$, there exist a state $\rho \in A \vee B$ such that the restriction, obtained via partial tracing over $\mathbf{H}_B$ (resp. $\mathbf{H}_A$) in the standard Hilbert space formulation, of that state to A (resp. B) is ρ_1 (resp. ρ_2). A condition, both necessary and sufficient, for this to be the case is that[6], for all $A \in A$ and $B \in B$, $\|AB\| = \|A\|\|B\|$.

Let us assume (i) i.e. that merely performing a measurement on one system cannot instantaneously affect our predictive capacities with respect to an independent 'subsystem' (in the above sense of independence). By saying that an operation Alice performs conveys no information to Bob, it is meant that the abstracted 'state of Bob's system' (in the authors phrasing) is left invariant by that operation. More precisely T, defined on $A \vee B$, conveys no information to Bob iff the restriction to B of T*, applied to the 'global' state

---

[1] *Ibid.*
[2] Despite its early recognition e.g. by Koopman 1931.
[3] Or with an objective 'property state' (Bub 1999, Chapter 1).
[4] Clifton *et al.* 2003.
[5] *Ibid.*
[6] Florig and Summers 1997.

$\rho \in \mathbb{A} \vee \mathbb{B}$, satisfies $[T^*(\rho)]_B = \rho_B$. If this is satisfied, then all expectation values for the outcomes of measurements of Bob's observables will not differ in any way after the operation from what they were before. Expressing T as a positive operator valued resolution of the identity in A, the condition is simply that for all observables $B \in \mathbb{B}$, T(B) = B, and this is shown by CBH to be equivalent to asserting that an effect on A, representing the measurement performed by Alice, commutes with any effect on Bob's side. Such commuting is thus equivalent to imposing the 'no instantaneous conveyance of information by measurement' constraint. It should be noted that the authors interpret the mutual commutation of A and B observables as an expression of *kinematical independence*, regardless of the fact that no consideration of a clearly mechanical nature is included in their premises – not least in the very definition of a C* state – nor in the course of their derivation.

Having established the existence and uniqueness of product states of $\mathbb{A} \vee \mathbb{B}$, such that $\rho(AB) = \rho(A)\rho(B)$ for all $A \in \mathbb{A}$ and $B \in \mathbb{B}$, where $\mathbb{A}$ and $\mathbb{B}$ are both C* and kinematically independent, CBH proceed to show that it suffices that the algebra of observables be abelian for a map to exist, which broadcasts (clones) an arbitrary pair of input mixed (pure) C*-defined states. Pairwise broadcasting (cloning) is thus always possible in a classical, hence abelian theory. In fact, abelianity is also a *necessary* condition for pairwise broadcasting to be possible. Since a quantum theory, in the author's C*-algebraic acceptation, is fundamentally non-abelian, broadcasting (cloning) is therefore precluded in a 'quantum' setting.

If binary preparations were restricted to product (pure) or separable (mixed) cases then, by properly acting upon spacelike separated 'subsystems', one could devise a bit commitment protocol that would be unconditionally secure. In other words, if their respective operations and cross-comparison of their locally gathered data can provide them only with evidence of ordinary ('classical') correlations, Alice cannot cheat by making a preparation she could later at will transform into either of two preparations associated with the committed bit, without Bob ever being able to detect the substitution. Assuming kinematical independence hence, according to (i), the absence of any instantaneous predictively significant repercussion on Bob's side of any measurement being performed by Alice on hers, and assuming also the non-commutativity of their respective algebras of observables, then enforcing the prohibition of unconditionally secure bit commitment[1] requires that entanglement should be a generic feature of binary preparations. The authors obviously understand this requirement as meaning that entanglement is endemic amongst pairs or n-uples of individual quantum systems, or that

[1] Behind Alice's capacity to cheat lies the basic fact, linked to non-commutativity, that different mixtures can be associated with the same density operator.

"spacelike separated systems occupy entangled states". Notice, however, that the aforementioned equivalence is derived on the basis of a notion of 'state' that bears no evident relationship to any form of 'occupation' associated with the possession of attributes by individual systems.

The main technical limitation of the CBH endeavour, which the authors duly acknowledge, is their inability to establish, within a C*-algebraic framework, the complete equivalence between a theory being 'quantum' and claiming that the three postulates hold. A proof that it is sufficient to allow entanglement in order to block the possibility of unconditionally secure bit commitment is still wanting. This is because, if the results of Mayers and Lo and Chau suggest it might be so, their proof hinges on properties of the elementary formalism of quantum theory – in particular, on the biorthogonal decomposition theorem, and a basic theorem[1] about the 'extractability' of any mixture from a given 'entangled state'. Deciding whether the implication holds in the C*-algebraic framework would require an appropriate generalisation of the latter theorem, which remains an open question.

CBH view their results as a significant step toward the goal of securing 'physically' compelling and solid foundations for the whole of quantum physics. In particular, they claim that "[t]he foundational significance of [their] derivation...is that quantum mechanics should be interpreted as a *principle theory*, where the principle at issue is information-theoretic."...Indeed, their principle-based characterisation of quantum theory is intended to "lend[s] credence to the idea that an information-theoretic point of view is the right response to adopt in relation to quantum theory." The suggested refoundation would be achieved by "substituting for the conceptually problematic mechanical perspective on quantum theory an information-theoretic perspective. That is, we are suggesting that quantum theory be viewed, not as first and foremost a mechanical theory of waves and particles, but as a theory about the possibilities and impossibilities of information transfer." In that respect, CBH make a parallel with STR, which is regarded as a paradigmatic principle theory :

> A relativistic theory is a theory with certain symmetry or invariance properties, defined in terms of a group of space-time transformations. Following Einstein, we understand this invariance to be a consequence of the fact that we live in a world in which natural processes are subject to certain constraints. A quantum theory is a theory in which the observables and states have a certain characteristic algebraic structure. Unlike relativity theory, quantum mechanics was born as a recipe[2] or algorithm for calculating the expectation values of observables measured by macroscopic measuring instruments.

---

[1] Hughston *et al.* 1993.

[2] This is a somewhat misleading statement: neither Heisenberg nor even Born (and obviously not Schrödinger) had a probabilistic interpretation in mind when they worked out the basic formalism of quantum mechanics (see Chapter 1).

> These expectation values (or probabilities of ranges of values of observables) cannot be reduced to probability distributions over the values of dynamical variables (or probability distributions over properties of the system). Analogously, one might imagine that the special theory of relativity was first formulated geometrically by Minkowski rather than Einstein, as an algorithm for relativistic kinematics and the Lorentz transformation, which is incompatible with the kinematics of Newtonian space-time. What differentiates the two cases is that Einstein's derivation provides an interpretation for relativity theory: a description of the conditions under which the theory would be true, in terms of certain principles that constrain the law-like behavior of physical systems. It is in this sense that our derivation of quantum theory from information-theoretic principles can be understood as an interpretation of quantum theory: the theory can now be seen as reflecting the constraints imposed on the theoretical representations of physical processes by these principles[1].

Thus a given type of algebraic structure is, according to CBH, *the* defining characteristic of a quantum theory. This is accordance with QIT practice and conceptions, but also with much modern work on the conceptual foundations of quantum mechanics[2]. Algebraic structure would be the reflection of certain definite constraints to which natural processes would be subject, and 'a' quantum theory would owe its particular structure to those constraints being distinctively information-theoretical. This might account for the long delay in our coming to terms with the more puzzling aspects of the basic quantum formalism. Information only came to the fore as a theoretical concept a good fifteen years[3] after the emergence of quantum mechanics, and engineering concerns with data processing were not extended to quantum-theoretical representations until the rise of QIT more than a half-century later.

But in what sense are the three constraints chosen by CBH information-theoretic? Although (i) is allegedly about the possibility of instantaneous information transfers 'between systems', no technical concept of information plays any part in the proof of equivalence with 'kinematical independence'. Given the fundamentally predictive significance of the C* notion of state, a more adequate expression of the constraint would be that performing a measurement on system A cannot (should not) by itself alter the manifold of probability distributions associated with measurements to be potentially performed on B alone. This is all there can be, realistically, to the kind of 'influence' the principle is intended to prohibit. If this is admitted, however, then it is somewhat inappropriate to think of the stated constraint as one to which natural processes, as opposed to an abstract agent's capacities of 'acting in the world', would be subject.

---

[1] Clifton *et al.* 2003.

[2] One of the three authors (Bub 1999) has been, for decades, advocating the importance of a logico-algebraic approach to those basic issues.

[3] Prominently through Norbert Wiener's cybernetics programme in the early 1940s (Dupuy 2000), then Shannon's 'theory of communication' (Shannon 1948, Shannon and Weaver 1949).

Of the three suggested postulates, the first is that which might appear to lend itself to the closest analogy with STR. This is deceptive, however, for the latter is not grounded on the recognition of an upper bound on the speed at which any physical influence can be propagated. In its original (1905) form, the theory only assumes that the speed of light is invariant in all references frames, an assumption that can in fact and should rather be dispensed with (see Section 3.1). Moreover, widespread though the reference to 'superluminal' influences or signalling is in the QIT literature, it must be repeated that neither at a technical nor at a conceptual level are considerations of relativity (Galilean or Einsteinian) involved. What is at stake here is the fact that two (n) observers who locally perform operations on 'parts' of a binary (n-ary) preparation exhibiting entanglement cannot transmit data to each other without actually exchanging some of the particles they measured, or performing further operations on their own particles on the basis of instructions supplied ('classically') by the other parties involved.

In its very content and formulation postulate (iii), which prohibits unconditionally secure bit commitment, is that of the three postulates which ties in most clearly with information-theoretical issues and concerns. Not so surprisingly perhaps, it is also the one to which the idea of a constraint to which *natural processes* should be subject is least applicable. All of the founding fathers, including Von Neumann (arguably the most likely to have responded with sympathy to the idea of bringing together physics and information theory[1]) would have been utterly perplexed with the suggestion to ground quantum physics, be in part, on the always open possibility for one party to cheat with total impunity whilst committing one bit to another. And this is not a mere matter of information-theoretical concerns not being 'in the air' when quantum mechanics emerged and its basic mathematical structure was unravelled. The urge to develop a new kind of mechanics arose as a response to the challenge of 'saving phenomena', whether these took the form of spectroscopic data, the black-body spectrum or the empirical values of the specific heats of solids. This cannot be simply brushed aside as a mere accident of history, that would have been both fortunate in its pointing (?) towards the need for a novel kind of framework, and rather unfortunate in its letting very little transpire of the actual necessities behind such a framework. It might well be, as CBH suggest, that quantum mechanics is not primarily about microscopic objects as such, be they conceived as particles, as waves or unlikely hybrids thereof. Indeed, the idea is put forward in chapter 3 of this dissertation that quantum mechanics can be seen to arise, albeit in an indirect manner, from the satisfaction of some rather abstract regulative constraints, encapsulated in Faraggi and Matone's 'equivalence principle' (Section 3.3).

[1] Bearing in mind that von Neumann was a co-originator of game theory, the theory of cellular automata and computer science.

On the other hand, it sounds prima facie rather implausible, to say the least, that given its relevance and its tremendously successful application to processes and physical structures on microscopic – typically (sub)atomic – scales, quantum mechanics, as a physical theory, should find the rationale of its efficiency in such broadly construed information-theoretical principles as suggested by CBH, Fuchs or others. Such endeavours should rather be thought of as alternative ways – possibly illuminating, more likely misleading – of pointing to essential differences between two kinds of *predictive frameworks*, associated with values 1 ('classical') and 2 ('quantal') of Hardy's r parameter (see Section 9.3).

End of Part B

10

Finale

It was argued (Part A, Chapter 3) that, given its status as a fundamental theoretical framework, one should reasonably expect quantum mechanics to be formulated as a principle theory. Indeed, Schrödinger's equation can be shown to follow, albeit in an indirect way, from the requirement that a Hamiltonian dynamical framework comply with a single ('Equivalence') principle. As discussed in Section 3.3, a principle account of quantum dynamics can be given on the basis of 'relativistic' requirements that relate to the suspension of any privilege that is classically granted to the rest frame. The Faraggi-Matone (FM) equivalence principle boils down to the statement that all mechanical systems are equivalent under coordinate transformations. In particular, there must exist a trivialising coordinate map that reduces any system to the free zero-energy case. A direct consequence of assuming the universal validity of the FM equivalence principle is an alternative (3.3.8) to the classical form of the stationary Hamilton-Jacobi equation. The (not simply Bohmian) 'quantum potential' whereby the 'quantum-mechanical' alternative (3.3.8 or QSHJE) to the Hamilton-Jacobi equation essentially differs from the latter is a direct consequence of postulating the universal existence of the trivialising map. That potential must be proportional to the Schwarzian derivative of the reduced action (a straightforward generalisation of Hamilton's characteristic function), which is the physically significant quantity in this principle-based refoundation of quantum mechanics. The new 'quantum reduced action' can never be constant, which guarantees the universal existence of the map. In contrast, the absence of any quantum potential in classical mechanics can be traced back to a suspiciously non-relativistic peculiarity of that frame in which a system is 'at rest'.

Schrödinger's equation turns out to be only derivative: it arises as a result of linearising the QSHJE. Consequently, and as a reflection of underlying symmetry, there is no one-to-one relationship between quantum reduced action and wave function. Therefore, a theoretical framework that, like quantum mechanics in its current acceptation, relies on Schrödinger's equation alone lacks the finer resolution, hence descriptive power, that a dynamical framework based upon (3.3.8) affords in principle. For the existence of the trivialising coordinate transformation, hence the non-existence of a constant reduced action to be secured, a general solution of the Schrödinger equation must have a bi-polar form (a real wave function must be the sum of a

polar expression and its complex conjugate). Whilst differing from Bohm's representation, this feature is in full agreement with Floyd's. Unlike Bohm's, Floyd's representation is based directly upon considering the solutions of the QSHJE (3.2.6-3.3.8), from which more general solutions of the Schrödinger equation can be subsequently derived. As Einstein had objected to Bohm[1], in the case of stationary systems and real wave functions (necessarily square-integrable in Bohm's account), the classical motion is not retrieved in the limit 'as Planck's constant tends to zero'. The objection is precisely met in the FM approach through the occurrence of a bipolar form, which reflects the impossibility for the reduced action to be constant if the FM equivalence principle holds.

Bearing in mind that it is the *stationary* quantum Hamilton-Jacobi equation that is derived from the FM equivalence principle, it is a condition, necessary and sufficient, for the Schwarzian derivative, hence for the quantum potential, to exist that the total energy E be such that the Schrödinger equation derived from (3.2.6) have E as an eigenvalue, or equivalently that the Schrödinger equation admit a solution that is a square-integrable function on the real line. On the other hand, that all solutions of the Schrödinger equation have to be square-integrable responds to no dynamical necessity. On the view here propounded, the usual and exclusive focus on square-integrability is justified only in so far as this property turns out to underlie the adequacy of a special kind of predictive scheme (SAQM). Such a scheme, which identifies with 'quantum theory' as it is commonly used and discussed, has proved to be most effective in its capacity to overcome in a pragmatic way the loss of descriptive power involved in switching from the solutions to the Q(S)HJE to those of the Schrödinger equation.

There remains the question of why FM equivalence, as reflected in the QSHJE, should go unnoticed unless the systems subject to theoretical treatment have typical actions that are much smaller than those that characterise 'classical' ones, and should (?) conversely be relevant and adequate to investigating (sub)molecular structure and behaviour (dimensional analysis shows that small action values typically coincide, albeit not necessarily, with microscopic scales). Why the proportionality factor in the non-homogeneous contribution to the QSHJE (FM quantum potential) should coincide with the empirically given Planck's constant is certainly no less, but no more mysterious than is the identification of the maximal permissible value of speed in STR with the empirical value of the speed of light. It is also intriguing, and perhaps quite significant though presently obscure, that both universal constants *c* and *h* should be so centrally involved in processes involving electromagnetic radiation, despite the fact that none of those

[1] See Holland 1993, p.243.

constants appears to be structurally bound by the nature of any particular interaction.

On account of the rather abstract character of the FM equivalence principle and of the technicalities involved in deriving the QSHJE, it is no wonder that no late nineteenth-century expert in analytical mechanics ever contemplated, be it as a purely intellectual exercise, any such amendment to the Hamilton-Jacobi theory. If anything, the endeavour presupposes relativistic concerns and insights that were not really 'in the air' before[1] 1905. Besides, if our hypothetical mathematician or physicist had, none the less, managed to derive the QSHJE, he could hardly have been aware of the underlying symmetries associated with the Schwarzian derivative and, beyond it, with the Legendre transformations. Last but not least, it is most unlikely he would ever have felt the urge, even with Hamilton's analogy between optics and mechanics in mind, to take the step of linearising the resulting 'quantum' alternative to the stationary Hamilton-Jacobi equation, thereby arriving at the 'Schrödinger' equation. If the occurrence in the quantum stationary Hamilton-Jacobi equation of an 'additional' potential-like term makes it inevitable to reassess basic tenets of classical mechanics (much as the replacement of the Galilean by the Lorentz transformation must lead to rethinking fundamental notions like inertia and energy), it is by no means clear a priori what the purpose and the significance of the Schrödinger equation might be, let alone that it should be crucial to an appropriate though probabilistic treatment of microphysical systems and processes. In hindsight, it is therefore not surprising that the actual emergence of quantum mechanics took an entirely different course. The breakthroughs of the first two decades of the twentieth century, from the resolution of the black-body or photo-electric puzzles to the Bohr-Sommerfeld model of the hydrogen atom, had given physicists no clue as to what kind of structural necessity the occurrence of Planck's quantum might be the manifestation of. A baffling duality of wave and particle aspects, which seemed from the outset to be part and parcel of the quantisation issue, was emphasised, rightly enough, although perhaps at the expense of another conceptually no less, and maybe far more important question: might the fact that Planck's constant has the same dimensions as classical action be a pointer toward its structural/physical significance? Faced with the urgency to come up with a unified account of the phenomenology of atomic physics, the matrix theorists on one side and Schrödinger on the other hit upon operator relations and differential equations that, quasi miraculously, led them to predictions that were empirically adequate. All of them failed to realise that Schrödinger's equation is in fact not the master equation of a theoretically

[1] The full significance of Einstein's *Zur Elektrodynamik...* paper was not immediately perceived (Miller 1998, 7.4).

well-grounded 'quantum' alternative to classical (Newtonian) mechanics. That the QSHJE can be derived *from* the Schrödinger equation through a formal substitution is by itself of little help to figure out what set of theoretical requirements could give rise to *that* equation (QSHJE) in the first place (this is arguably the most basic shortcoming of Bohm's approach). For the sake of rapid progress in the development of atomic physics, it was most fortunate that mathematical properties of solutions of the Schrödinger equation should lend themselves to the formulation of a most effective probabilistic algorithm. Again, it so happens that square-integrable solutions of the Schrödinger equation have the requisite mathematical features for setting up a consistent linear scheme, the aim of which is to compute the probabilities of observational outcomes. It took decades, however, for this to be fully recognised, and for it to be realised that such a predictive scheme owes its characteristic structure to its adequacy for the calculation of probabilities of context-dependent occurrences. The best generalisation to date of such a scheme is a Hardy r = 2 theory[1]. The linearity of both r = 1 and r = 2 Hardy frameworks, hence that of the SAQM, certainly reflects, at least in part, the linearity that is an inherent feature of averaging. However, this is not to say that linearity in the Schrödinger equation itself actually is a consequence of its yielding an adequate algorithm for computing expectation values. This would be reversing both theoretical priorities and historical facts. Rather, it is the case that the Schrödinger equation, the linearity of which arises from the very mathematical procedure whereby it is derived from the QHJE, lends itself to setting up a predictive algorithm that is isomorphic to a projector-based instantiation of a Hardy (r = 2)-type scheme. The implementation of such an algorithm is made possible by Hilbert space structure, which is itself abstracted from the existence of square-integrable solutions to the Schrödinger equation. At any rate, dynamics does – and must – come first, be it in the conceptually obscure embodiment of the Schrödinger equation (or 'Heisenberg picture'). The emergence of a most efficient probabilistic scheme for prediction should thus essentially be regarded as a practically welcome, but conceptually misleading by-product of a more basic and well-motivated alternative to classical Hamilton-Jacobi theory.

Hilbert space structure, which reflects the truly 'quantal' dynamics regulated by the QHJE only in an indirect manner, constrains the form of physical quantities via the linear representation of the relevant groups[2]. The linear operators that represent fundamental observables, with their distinctively 'quantum-mechanical' commutator structure, arise as a response to requirements of invariance under time, space and inertial transformations. A Hilbert space-based scheme for the calculation of probabilities of measurement outcomes is

[1] Section 9.3.
[2] Section 2.2.

optimal given the ensuing limitations to the mutual compatibility of such observables – in other words, the scheme has no 'surplus structure'. Indeed, "if all observables were compatible, not only would the Hilbert space have, as it were, some surplus capacity for the representation of observables, but a number of rays would represent the same pure[1] [case]." The dynamics regulated by the QHJE exerts fundamental constraints on the kind of linear vector space structure on the basis of which a consistent and effective predictive scheme can be set up (via Schrödinger's equation). These include the necessity for Hilbert space to be defined over the field of complex numbers[2] (such necessity can be traced back to symmetry associated with Legendre duality and time evolution; see Section 3.3), which contributes in a crucial manner to the optimality of the predictive scheme.

If the mutual compatibility of experimentally accessible quantities is restricted to non-equivalent classes, all experimental propositions cannot be coordinated in a unified way within a boolean framework, isomorphic to the algebra of set-theoretical operations. Instead, the validity of boolean logic and the Kolmogorovity of probability can be maintained only within precisely circumscribed contexts. Predictive consistency across non-equivalent contexts then calls for a metacontextual logic with a lattice structure that is orthocomplemented and non-distributive. This point was emphasised several decades ago by P. Heelan[3], who tentatively argued that "the locus of nonclassical logic in quantum mechanics is in the plane in which physical contexts are related to one another, and not, as all writers have hitherto held, in the plane of single quantum-mechanical events[4]". According to Heelan, "the proper subject matter of so-called quantum logic would be the manifold of experimental contexts in which it is relevant to use one linguistic or conceptual framework rather than another[5]" (*ibid.*). Its characteristic (orthocomplemented and *non*-distributive) lattice structure would then consist in a partial ordering of a set of experimental languages that are pairwise incompatible, in the sense that each one corresponds to a distinct choice of basis for spanning the linear subspace the ket ascribed to the current preparation belongs to (each basis choice picks off a definite context, hence a definite class of meaningful, testable experimental propositions). Thus, the need for 'quantum' logic would arise because limits to the compatibility of quantities in Hilbert space quantum

---

[1] Hughes 1989, p.112.

[2] Attempts to cast quantum theory in real Hilbert space (Stückelberg 1960 ; Mackey 1963) have shown that preserving the full and empirically adequate predictive potential of the SAQM requires e.g. introducing an *ad hoc* operator **J**, which commutes with all the operator representatives of observables, and such that $\mathbf{J}^2 = -\mathbf{I}$. Since the latter relation is just another representation of the imaginary number *i*, this is tantamount to reverting to the complex formulation.

[3] Heelan 1970a,1970b.

[4] Heelan 1970b.

[5] *Ibid.*

theory enforce a metacontext-language (used to coordinate propositions that pertain to context-related languages) that is precisely structured as a non-distributive lattice.

A predictive framework that is made possible via the square-integrability of solutions to the Schrödinger equation will typically rely on the identification of a context-neutral mathematical entity or predictor (Hilbert space vector, density matrix) used to characterise the predictive yield of a given preparation upon the selection of an appropriate context. An adequate predictive scheme must obviously preserve the validity of the ordinary (Kolmogorovian) probability calculus within any given context, but the form of its basic rules also reflects a characteristic meta-context lattice structure. Although neither Hardy's axioms nor Destouches's more rudimentary scheme[1] make any explicit reference to context-dependence, it is satisfying to think of the statistical algorithm of quantum theory – i.e. of quantum theory *tout court* as it is usually debated on or conceived of in QIT – as coordinating probabilistic valuations at a meta-context level, in other words as an algorithm that functions both *within* and *across* definite contexts.

Tensor product composition allows the predictive formalism to be applicable to 'fragmented' or *n*-ary preparations. Whether any such preparation can be empirically realised, and what it amounts to physically, e.g. in terms of constituent subsystems and properties thereof, falls beyond the compass of the SAQM. In any case, comparing locally gathered data may exhibit correlations that cannot be accounted for on the basis of a simple-minded picture based on common causes. Owing to the purely predictive scope of the scheme, this means nothing but that some commonsense expectations associated with classical probability calculus (Hardy r = 1 theory) need not apply to a 'non-classical' predictive scheme (r = 2 theory). In particular, the fact that correlations are predicted, which cannot be interpreted as originating in any sort of preordained 'instruction set' does not imply the existence of non-local influences or any other kind of spooky links between distant localised entities[2]. Whilst it is legitimate to inquire about the way mutual relationships between parts of a composite system should be treated in a quantum dynamical framework based on the QHJE, and whether this should call for holism or non-locality, those are matters to be addressed on the basis of *that* equation and of its solutions.

The question of non-locality primarily arose in discussions of the Einstein-Podolsky-Rosen gedankenexperiment[3], and the violation of Bell's inequalities[4]

---

[1] Section 5.3.

[2] Mermin 1999 goes so far as to dismiss all claims of non-locality in the quantum setting as mere "fashion at a distance".

[3] Einstein *et al.* 1935; reprinted in Wheeler and Zurek 1983.

[4] Bell 1964,1987.

by statistics of measurements performed on 'entangled systems'. In a simple version[1] of the EPR experiment, diatomic molecules with zero total angular (spin) momentum split into two identical atoms I and II whose spins are opposite. The resulting two beams travel in opposite directions and interact separately with Stern-Gerlach magnets: I-atoms 'on the left' (L), II-atoms 'on the right' (R). All predictions relating to observations performed, with that particular set-up, on I and II atoms, including all correlations between results of such observations, can be calculated using the rules of quantum theory given a ket of the form (8.1.1).

The direction of the magnetic field and its gradient is the only characteristic of each of the Stern-Gerlach (SG) magnets that is deemed relevant to spin analysis. Let that direction be given by the vector $\vec{B}_L$ for the SG device on the left, and by $\vec{B}_R$ for that on the right. Let us further assume each particle behaves, for all the purposes of the experiment, like a microscopic spinning top. A definite and separate spin attribute for each of the atoms of any given pair would then be tantamount to the specification of a 'preferred' spatial direction. If it is assumed that the spin orientation of an atom is randomly selected at the time when the I-II pair is created, the relevant parameter λ in Bell's analysis will be an angular value, treated as a random variable with a uniform distribution over the interval $[0,2\pi]$. Any I-II correlation is then imputable to a rigid symmetrical relationship between values of λ_I and λ_{II}, irrespective of the orientation of the SG magnets.

$\vec{B}_R$ can be chosen to provide a reference direction for angular deviations of the I and II beams. If the orientations of the analysers are parallel, the I and II deviations will be observed to be systematically opposite. On the other hand, if $\vec{B}_R$ and $\vec{B}_L$ make an angle θ, then this very sketchy model leads one to expect 'correlation errors', i.e. failures of such oppositeness, which should occur with a probability that is proportional to θ. Owing to the hypothesis of angular equiprobability, the probability of any such error is $p_e(\theta)=\frac{|\theta|}{\pi}$ and the correlation is $E(\theta)=p_e(\theta)-(1-p_e(\theta))=\frac{2|\theta|}{\pi}-1$. The latter satisfies Bell's inequalities

$$|E(\theta_{11})+E(\theta_{12})\pm E(\theta_{22})\mp E(\theta_{21})|\le 2\ ,$$

[1] Bohm 1951.

where the θ_{ij} denote relative orientations of the SG analysers[1]. By contrast, the prediction of quantum theory is $p_e(\theta) = \frac{1-\cos\theta}{2} = \sin^2\frac{\theta}{2}$ and $E(\theta) = -\cos\theta$, for which there are choices of the θ's that violate the inequalities.

A convenient picture of the difference between the two kinds of distribution is given by a unit circle with a preferred diameter (D), the direction of which is chosen parallel to that of the field gradient of one of the SG devices (e.g., that on the right). A point M on the circle can be taken to correspond to a given and hypothetically predetermined orientation of spin for the corresponding particle (e.g. II on the right). If the point N is the result of an orthogonal projection of M onto (D), then if M falls within one of the two angular sectors of amplitude θ for which there may arise correlation errors, the corresponding N will fall within a line segment of (D), the length of which is $1 - \cos\theta$. Let us now transfer the randomness assumption from angular values to their cosines, i.e. from the M points to the N points. A uniform distribution, which corresponds to equiprobability along D, then takes the form that quantum theory predicts[2]: $p_e(\theta) = \frac{2(1-\cos\theta)}{4} = \sin^2\frac{\theta}{2}$.

A probability p_e that is proportional to θ reflects the initial assumption that the angular λ parameters are randomly distributed at the time when the I-II pairs are created, and that any possible future interaction with a SG device is totally irrelevant to the ensuing statistics. In contrast, the analysers force upon the experiment a preferred choice of directions[3]. Thus, the rationale for a violation of Bell-type inequalities appears to come down to whether the probability law is deemed an (indirect) expression of respective physical attributes of I and II, and should therefore only be dependent on the local circumstances of the creation of the I-II pairs; or whether prediction, in that particular set-up, should only hinge upon the relative orientation of two suitably positioned analysers. At any rate, there is no logical necessity in requiring that the correct probability distribution should be in agreement with a preconception of I and II as anything remotely like microscopic marbles produced with definite and randomly selected spinning axes. Besides, the first horn of the dilemma overlooks the non-isotropy of the set-up, which is, predictively speaking, its

---

[1] For example, the maximum value of 2 obtains for $\theta_{12} = \theta_{21} = \theta_{11} + \frac{\pi}{2}$ and $\theta_{22} = \theta_{11} + \pi$.

[2] Although a given N may correspond to two distinct points on the circumference, ambiguity can be removed by distinguishing the two semicircles (D) separates and by counting the diameter once for each semicircle; this accounts for the number 4 in the denominator of the following expression.

[3] In either case, it should not matter whether the direction that is chosen as a reference is that of the field gradient of one of the SG analysers rather than the other. In other words, the probability law ought to, and does indeed depend only on the relative orientation of the analysers.

most significant characteristic Should not, on the contrary, the probability distribution reflect the symmetries – or lack thereof – of the experimental arrangement, which in this case boils down to relative orientations of the SG analysers? If so, then if one can speak at all of any influence on II of one's measuring I, it is only in the sense that those relevant characteristics of the set-up should determine specific correlations, which comparing data obtained separately on I and II would exhibit. In any case, such dependence of prediction on basic traits of the set-up should not lead one astray into contemplating the existence of spooky links between the two atoms of any given pair[1].

Let us stress it again: the effectiveness of the predictive rules of Section 4.2 – or some generalisation thereof – must be grounded in a proper dynamical account, if they are not to be vacuous, or their adequacy incomprehensible. This must be borne in mind whilst assessing anti-realist construals of quantum theory, which may draw on such material as that covered in Part B of this dissertation. An especially significant example is provided by the theses propounded by Michel Bitbol. In a number of thought-provoking books and papers[2], Bitbol has argued that a correct philosophical appraisal of quantum physics should have the beneficial – indeed, almost therapeutic – effect of dispelling a major illusion: that a phenomenon can always be in principle detached from the very conditions that make its occurrence possible. Bitbol's stance is clearly neo-kantian. However, his more recent writings attempt to minimise the idealistic flavour of a 'transcendental reduction' via emphasising scientific practice and cognitive-operational preconditions to its effectiveness. Another acknowledged influence is that of Bohr and Heisenberg. Bitbol reads them as both striving, perhaps in a rather inadequate fashion, to articulate the view that such phenomena as encountered in atomic physics are necessarily and indissolubly co-determined by the experimental conditions of their manifestation. Provided our anthropological situation of cognisers and experimenters is properly accounted for, there would be "no need to further explain" the efficiency of quantum rules "by their ability to reflect in their structure the backbone of nature[3]". Rather, Bitbol contends, "the basic formalism of quantum mechanics can effortlessly be construed as a structural presupposition of any activity of production and unified anticipation of mutually

---

[1] This might well be what Bohr had meant in his notoriously obscure reply to Einstein, Podolsky and Rosen (Bohr 1935; reprinted in Wheeler and Zurek 1983); Halvorson and Clifton 2001.

[2] Bitbol 1996,1998,2002.

[3] Bitbol 1998. By contrast, asserting that "self-existent objects are what justify the intentional attitudes" would be "very imprudent", and "[the] project of ontologizing certain theoretical entities appears a mere attempt at hypostatising the major invariants of those activities." (*ibid.*)

incompatible contextual phenomena[1]." This belated recognition of the aim and structure of quantum theory would eventually

> "undermine the pictures so cherished by supporters of the ontological (disengaged) outlook... by showing that the predictive success of some of our most general scientific theories can be ascribed, to a large extent, to the circumstance that they formalize the minimal requirements of any prediction of the outcomes of our activity, be it gestural or experimental. The very structure of these theories is seen to embody the performative structure of the experimental undertaking[2]."

Bitbol bases his claims entirely on (i) Destouches's reconstruction of the basic formalism of quantum theory (SAQM) and (ii) Heelan's tentative appraisal of quantum logic[3]. These prompt an outright denial of the physical reality of Schrödinger's 'Ψ waves': the sole purpose of Ψ functions or kets is to provide abstract predictors (in Destouches's sense, which Bitbol makes his own without further criticism), the mathematical properties of which would only reflect their invariance under changes of experimental situations (contexts). What remains of 'complementary' corpuscular aspects is a metaphorical way of referring to series of discretely distributed events as they occur under given experimental conditions. Thus one could say that quantum 'waves' are physically vacuous appearances that manifest the *intercontextual* effectiveness of a particular kind of predictive scheme, whereas the appearance of particles is a manner of describing, in a rather inappropriate classical mode, the discreteness of *intracontextual* occurrences. When all is said and done, the *inter* vs. *intra* distinction is all that should be left of Bohr's half-baked idea of complementarity.

Given Bitbol's rejection of any kind of disturbance based view, it is somewhat unclear why it should have taken the physicists' confrontation with new – typically microscopic – domains of experience for them to realise that our means and procedures of investigation cannot be neutralised in all circumstances, and that they may in fact be constitutive of the very objects of investigation. On the other hand, it hardly comes as a surprise that, rather than seize the opportunity to re-assess preconditions of our cognitive-empirical activities within a world-as-experienced, many should have remained caught within the illusion that science can, in all circumstances, afford a disengaged outlook and have access, in a 'strongly objective[4]' way, to truths about a real-world-out-there.

Should we then consider, as Bitbol invites us, "quantum mechanics as a general *technology of mesoscopic action and experimentation*, or as a

---

[1] Bitbol 2002.
[2] Bitbol 1998.
[3] Heelan 1970a,1970b.
[4] d'Espagnat 1995.

dialectical relation between situated phenomena and predictive invariants[1]"? It is the object of Part B of the present thesis to show that the SAQM functions as a context-sensitive kind of predictive formalism, in which the actual form of the relevant physical quantities is determined by symmetries. This suffices to account for the basic structure and the effectiveness of quantum theory, and "automatically defuses major paradoxes[2]". Is this, however, all that one should reasonably expect from a fundamental physical theory such as quantum mechanics purports to be, and is generally regarded as? Bitbol's answer is unambiguous: "[t]he only thing a physical theory does, and the only thing it has to do, is to embed documented actualities in a (deterministic or statistical) framework, and to use this framework to anticipate, to a certain extent, what will occur under well-defined experimental circumstances[3]". Science can and should be concerned only with the web of interrelationships woven by carefully conducted experimentation/observation and the relevant modes of communication[4], in contrast to classical ('pre-critical') views that "systematically favour a disengaged outlook, even though their very undertaking is grounded on the presuppositions of an engaged activity[5]." Duhem's remark[6] that ontological pursuits in physical science can be motivated only by 'reasons of the heart', which neither analytical reason nor experience can prove legitimate, is as cogent today as it was in the first decade of the twentieth century. All the same, and whatever the seductions of the transcendental programme, it sounds rather implausible that physicists between 1900 and 1926 did, through the finiteness of Planck's constant, unwittingly stumble upon nothing but the empirical trace (?) of regulative constraints their cognitive and experimental activity cannot but comply with.

Bitbol also fails to address in any convincing manner such issues as the emergence in the long run, in a given 'two slits open, no which-way detection attempted' experimental situation, of an interference pattern that seems to indicate that some genuinely physical and, in a sense, 'wavelike' process is going on. Whilst 'wavelike' cancellation or reinforcement of terms in the calculation of probabilities need not raise any qualms about its physicality or lack thereof, the actual build-up of fringe patterns in a two-slit or other kinds of experiments cries out for an explanation (which standard quantum mechanics notoriously fails to provide). If anything, it sounds almost incongruous that a basic mathematical trait associated with the flexible use of a predictive scheme 'across' non-equivalent contexts should be underwritten by specific

---

[1] Bitbol 2002.
[2] *Ibid.*
[3] Bitbol 1998.
[4] Our most successful theoretical frameworks are to be seen as "embodiments of the necessary pre-conditions of a wide class of activities of seeking and predicting" (Bitbol 1998).
[5] Bitbol 1998.
[6] Duhem 1954.

patterns exhibited on photographic plates. This is precisely the sort of pitfalls the principle (re)foundation of quantum mechanics suggested in Part A of this thesis avoids. For then, although the basic predictive rules of quantum theory can thoroughly be justified without making any reference to an underlying ontology of physical systems, just as important is the realisation that, while the current predictive scheme owes its structure to properties of a certain class of solutions of the Schrödinger equation, the latter is (merely) a by-product of a principle-based dynamical framework that is regulated by the QHJE. One cannot therefore, as Bitbol – or QIT-inspired thinkers whose own motivations are extremely different – would invite us to do, so readily dispense with any claim of 'physicality' in regard of the predictive efficiency and conceptual significance of *quantum mechanics.*

On the principle view advocated in Chapter 3, what is called for is a radical shift of attention from the Schrödinger equation and toward the Q(S)HJE. Investigating solutions of the latter equation cannot be bypassed if fundamental aspects of 'quantum behaviour' are to be elucidated – one example among thousands is the vexed problem of accounting in a fully intelligible manner for the fine details of the actual emergence of an interference pattern in a two-slit experiment, beyond the mere statistical distribution that the SAQM predicts for a given choice of experimental set-up, and without resorting to such baffling 'explanations' as 'self-interference' of individual particles. Our hopes and ambitions are most likely to be tempered by analytical difficulties, although numerical methods should be very helpful where the techniques of calculus are powerless to reach. However difficult the task might prove to be, the endeavour cannot and hopefully will not be indefinitely postponed, for only then will our understanding of quantum dynamics and our appraisal of conceptual issues measure up to our predictive abilities.

Bibliography*

Alber G., Beth T., Horodecki M., P. and R. 2001. *Quantum information – an introduction to basic theoretical concepts and experiments*, Springer Tracts in Modern Physics, Springer Verlag, Berlin.

Albert D.Z. and Loewer B. 1988. *Interpreting the Many-Worlds Interpretation*, Synthese **77**, 195-213.

Aspect A., Grangier P. and Roger G. 1982. 'Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment : A new violation of Bell's inequalities', *Phys. Rev. Lett.* **49**, 91-94 ; 'Experimental test of Bell's inequalities using time-varying analyzers', *Phys. Rev. Lett.* **49**, 1804-1807.

Barnum H., Caves C.M., Fuchs C.A., Jozsa R. and Schumacher B. 1996. 'Noncommuting mixed states cannot be broadcast', *Phys. Rev. Lett.* **76**, 2818-2821 [quant-ph/9511010].

Barrow J., Davies P. and Harper C. (eds.) 2003. *Science and Ultimate Reality*, Cambridge University Press.

Bell J.S. 1964. *Physics* **1**, 195; reprinted in Bell 1987.

Bell. J.S. 1987. *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press.

Bennett C.H. 1992. 'Quantum cryptography using any two nonorthogonal states', *Phys. Rev. Lett.* **68**, No.21, 3121-3124.

Bennett C.H. and Brassard G. 1984. 'Quantum Cryptography: Public Key Distribution and Coin Tossing', in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing*, pp. 175-179.

Bennett C.H. and Wiesner S. 1992. 'Communication via one and two-particle operators on Einstein-Podolsky-Rosen states', *Phys. Rev. Lett.* **69**, 2881.

Bennett C.H., Brassard G. , Crépeau C., Jósza R., Peres A. and Wootters W.K. 1993. 'Teleporting an unknown quantum state via dual classical and Einstein-Podolski-Rosen channels', *Phys. Rev. Lett.* **70**, 1895-1899.

Bennett C.H., DiVincenzo D.P., Fuchs C., Mor T., Rains E., Shor P.W., Smolin J.A., Wootters W.K. 1999. 'Quantum Nonlocality without Entanglement', *Phys. Rev.* **A**, 1070 [quant-ph/9804053].

Bernstein J. 1991. *Quantum Profiles*, Princeton University Press, Princeton NJ.

Bhatia R. 1997. *Matrix Analysis*, Springer-Verlag, New York.

Bhattacharyya A. 1946. 'On the measure of divergence between two multinomial populations', *Sankhyâ, The Indian Journal of Statistics* **7**, 401-406.

Birkhoff G. and von Neumann J. 1936. 'The Logic of Quantum Mechanics', *Annals of Mathematics* **37**, 823-843. [Reprinted in Hooker 1975, Vol. 1].

*Quantum physics e-prints, listed as quant-ph/yymmnnn, are available at http://xxx.arXiv.org (mirror site: http://xxx.soton.ac.uk).

Bitbol M. 1996. *Mécanique quantique – Une introduction philosophique*, Flammarion, Paris.

Bitbol M. 1998. 'Some steps toward a transcendental deduction of quantum mechanics', *Philosophia Naturalis* **35**, 253-280.

Bitbol M. 2002. 'Science as if situation mattered', *Phenomenology and Cognitive Science*, **1**, 181-224. Available at http://philsci-archive.pitt.edu .

Blakemore C. and Greenfield S. (eds.) 1987. *Mindwaves*, Basil Blackwell, Oxford.

Bohm D. 1952. 'A suggested interpretation of the quantum theory in terms of "hidden variables"', *Phys. Rev.* **89**, 166-193.

Bohm D. 1983. *Wholeness and the Implicate Order*, Ark paperbacks, London.

Bohm D. and Hiley B.J. 1993. *The Undivided Universe*, Routledge, London.

Bohr N. 1928. *Atti del Congresso Internazionale dei Fisici,* Zanichelli, Bologna, vol.2, pp.565-588.

Bohr, N. 1935. 'Can quantum-mechanical description of physical reality be considered complete?'. *Phys. Rev.* 48, 696-702. Reprinted in Wheeler and Zurek 1983.

Boschi D., Branca S., DeMartini F., Hardy L. and Popescu S. 1998. 'Experimental Realization of Teleporting an Unknown Pure Quantum State via Dual Classical and Einstein-Podolski-Rosen Channels', *Phys. Rev. Lett.* 80, 1121-1125.

Bose S. , Vedral V. and Knight P.L. 1998. 'A Multiparticle generalisation of entanglement swapping', *Phys. Rev.* A **57**, 822 [quant-ph/9708004].

Bose S. , Vedral V. and Knight P.L. 1999. 'Purification via entanglement swapping and conserved entanglement', *Phys.Rev.* A **60,** 194 [quant-ph/9812013].

Bouwmeester D., Pan J.W., Weinfurter H. and A. Zeilinger 1998. Phys. Rev. Lett. **80**, 3891.

Braunstein S.L. , Mann A. and Revzen M. 1992. Phys. Rev. Lett. **68**, 3259.

de Broglie 1928. 'La nouvelle dynamique des quanta', in *Electrons et Photons, Rapports et discussions du cinquième Conseil de Physique Solvay*, Paris, Gauthier-Villars, pp. 105-141.

de Broglie L. 1956. *Nouvelles perspectives en microphysique.* Albin Michel, Paris.

Brown H.R. and Redhead M.L.G. 1981. *A Critique of the Disturbance Theory of Indeterminacy in Quantum Mechanics*, Foundations of Physics **11**, 1-20.

Bub J. 1999. *Interpreting the Quantum World*, Cambridge UP (1st ed. 1997).

Bub J. 2000. 'Quantum Mechanics as a Principle Theory', *Studies in History and Philosophy of Modern Physics* **31B** No.1, 75-94.

Busch P., Grabowski M. and Lahti P. 1995. *Operational Quantum Physics*, Springer-Verlag, Berlin.

Busch P.1996. 'Is the quantum state (an) observable?', in Cohen and Stachel 1996 [quant-ph/9604014]

Busch P. 2003. 'Quantum states and generalized observables: a simple proof of Gleason's theorem', *Phys. Rev. Lett.* 91, 120403 [quant-ph/9909073].

Butterfield J. 1989. 'The Hole Truth', *British Journal for the Philosophy of Science* **40**, 1-28.

Butterfield J. and Placek T. (eds.) 2002 . *Non-locality and Modality*: Proceedings of the NATO Advanced Research Workshop on *Modality, Probability, and Bell's Theorems*, Cracow, Poland, August 19-23, 2001 (NATO Science Series II: Mathematics, Physics & Chemistry), Kluwer

Caticha A. 1998. 'Consistency, amplitudes, and probabilities in quantum theory', *Phys. Rev.* A**57**, 1572. [quant-ph/9804012]

Caticha A. 2000. 'Insufficient reason and entropy in quantum theory', *Found. Phys.* **30**, 227-251 [quant-ph/9810074].

Caves C.M. and Fuchs C.A. 1996. 'Quantum Information: How Much Information in a State Vector?', in Mann and Revzen 1996.

Caves C.M., Fuchs C.A. and Schack R. 2002. 'Unknown Quantum States: The Quantum de Finetti Representation', *J. Math. Phys.* **43**, 4537 [quant-ph/0104088].

Caves C.M., Fuchs C.A., Manne K. and Renes J.M. 2004. Gleason-Type Derivations of the Quantum Probability Rule for Generalized Measurements, *Found. Phys.* **34**, 193 [quant-ph/0306179].

Chernoff H. 1952. 'A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations', *The Annals of Mathematical Statistics* **23**, 493-507.

Clifton R. 2001. 'The Subtleties of Entanglement and its Role in Quantum Information Theory', preprint (http://philsci-archive.pitt.edu)

Clifton R., Bub J. and Halvorson H. 2003. 'Characterizing Quantum Theory in terms of Information-Theoretic Constraints', *Foundations of Physics* **33**, 1561-1591 [quant-ph/0211089].

Cohen L. 1988. 'Rules of Probability in Quantum Mechanics', *Found. Phys.* **18** (10), 983-998.

Cohen R.S. and Stachel J. (eds) 1996. *Experimental Metaphysics – Quantum Mechanical Studies in Honor of Abner Shimony*, D.Reidel, Dordrecht.

Cover T.M. and Thomas J.A. 1991. *Elements of Information Theory.* Wiley Series in Telecommunications, John Wiley & Sons, New York.

Cox R.T. 1946. Am. J. Phys. **14**, 1.

Darrigol O. 1986. 'The Origin of Quantized Matter Waves', *Historical Studies in the Physical and Biological Sciences* **16**, 197-253.

Darrigol O. 1992. *From c-numbers to q-numbers: The classical analogy in the history of quantum theory*, University of California Press, Berkeley.

Davies P.C.W. 1974. *The Physics of Time Asymmetry*, University of California Press, Berkeley.

Davies, E.B. 1976. *Quantum Theory of Open Systems*, Academic Press, New York.

Davies E.B. 1978. 'Information and Quantum Measurement', *IEEE Trans. Inf. Theory* **IT-24**, 596-599.

Deltete R. and Guy R. 1990. 'Einstein's opposition to the quantum theory', *Am. J. Phys.* **58** (7), 673-688.

Destouches J.L. 1939. *Physique moderne et philosophie*, Hermann, Paris.

Destouches J.-L. 1941. *Corpuscules et systèmes de corpuscules*, Gauthier-Villars, Paris.

Destouches J.-L. 1942. *Principes fondamentaux de Physique théorique*, Hermann, Paris.

Destouches-Février P. 1946, 1951, 1956. See Février P. 1946, 1951, 1956.

Deutsch D. 1997. *The Fabric of Reality*, Viking Penguin Publishers, London.

Deutsch D. 2003. 'It From Qubit', in Barrow *et al.* 2003.

De Witt B.S. and Graham N. (eds) 1973. *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton University Press.

Dieks D. 1982. 'Communication by EPR devices', *Phys. Lett.* A **92**, 271-272.

Dieks D. 1988. *Phys. Let.* A **126**, 303.

Dirac P.A.M. 1958. *The Principles of Quantum Mechanics* [1930] 4th ed., Clarendon Press, Oxford.

DiVincenzo D.P. and Peres A. 1996. 'Quantum codewords contradict local realism', *Phys. Rev.* A**55** (1997) 4089 [quant-ph/9611011].

Duhem P. 1954. *The aim and structure of physical theory,* Princeton/Oxford University Press [transl. by P.P. Wiener of "La théorie physique : son objet, sa structure", 1906].

Dunn K.A. 1981. 'Poincaré group as reflections in straight lines', *Am. J. Phys.* **49**(1), 52-55.

Dupuy J.P. 2000. *The Mechanization of the Mind*, Princeton University Press, Princeton/ Oxford.

Earman J. 1995. *Bangs, Crunches, Whimpers and Shrieks: singularities and acausalities in relativistic spacetimes*, Oxford University Press, New York.

Ekert A.K. and Jozsa R. 1998. *Quantum Algorithms: Entanglement Enhanced Information Processing*, Phil. Trans. Roy. Soc. (Lond.), Proceedings of Royal Society Discussion Meeting *Quantum Computation: Theory and Experiment*, held in November 1997[quant-ph/9803072]

Einstein A.1919. 'Time, Space and Gravitation', *Optician*, *The British Optical Journal* **58**, 187-188. Also published under the title 'What is the Theory of Relativity?' in the November 28, 1919 issue of *The Times*, London. Reprinted in Einstein 1954.

Einstein A. 1954. *Ideas and Opinions*, Bonanza Books, New York.

Einstein A. 1989. *Oeuvres choisies*, 1. *Quanta.* Seuil/CNRS, Paris.

Einstein A. , Podolski B. and Rosen N. 1935. 'Can Quantum-Mechanical Description of Physical Reality be Considered Complete ?', *Phys. Rev.* **47**, 777-780. Reprinted in Wheeler and Zurek 1983.

d'Espagnat B. 1976. *Conceptual Foundations of Quantum Mechanics*, 2nd ed., Addison-Wesley, Reading MA.

d'Espagnat B. 1995. *Veiled Reality. An Analysis of Present-Day Quantum Mechanical Concepts*, Addison-Wesley, Reading MA.

Everett III H. 1957. '"Relative state" formulation of quantum mechanics', *Rev. Mod. Phys.* **29**, 454-462.

Everett III H. 1973. *The Theory of the Universal Wave Function*, in De Witt and Graham 1973 , pp. 3-137.

Fano G. 1971. *Mathematical Methods of Quantum Mechanics*, McGraw Hill, New York.

Faraggi A.E. and Matone M. 2000. 'The Equivalence Postulate of Quantum Mechanics', *Int. J. Mod. Phys.* **A15**, 1869-2017 [hep-th/9809127]

Février P. 1946. 'Signification profonde du principe de décomposition spectrale', *Comptes Rendus de l'Académie des Sciences* **222**, 866-868.

Février P. 1951. *La structure des théories physiques*, Presses Universitaires de France, Paris.

Février P. 1956. *L'interprétation physique de la mécanique ondulatoire et des théories quantiques*, Gauthier-Villars, Paris.

Feynman R.P., Leighton R. and Sands M. 1965. *The Feynman Lectures on Physics – Quantum Mechanics*, Addison-Wesley, Reading MA.

Feynman R.P. and Hibbs A.R. 1965. *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.

Feynman R.P. 1982. 'Simulating Physics with Computers', *Int. J. Theor. Phys.* **21**, Nos. 6/7, 467-488.

Feynman R. P. 1986. *Found. Phys.* **16**, 507.

Fine A. 1986. *The Shaky game : Einstein, Realism and the Quantum Theory*, University of Chicago Press.

de Finetti B. 1964. 'Foresight: Its Logical Laws, its Subjective Sources', in Kyburg and Smokler 1964 (pp.93-159) [translation of 'La Prévision, ses lois logiques, ses sources subjectives', *Annales de L'Institut Henri Poincaré* 7 (1937)]

Finkelstein D. 1962. 'The Logic of Quantum Physics', Trans. New York Acad. Sci. 25, 621-637.

Fisher R.A. 1922. 'On the dominance ratio', *Proc. Roy. Soc. Edimburgh* **42**, 321-341.

Fisher R.A. 1956. *Statistical Methods and Statistical Inference*, Hafner, New York.

Florig M and Summers S. 1997. 'On the Statistical Independence of Algebras of Observables', *Journal of Mathematical Physics* **38**, 1318-1328.

Floyd E.R. 1996. 'Where and Why the Generalized Hamilton-Jacobi Representation Describes Microstates of the Schrödinger Wave Function', *Found. Phys. Lett.* **9**, 489-497 [quant-ph/9707051]

Floyd E.R. 2000. 'Extended Version of "The Philosophy of the Trajectory Representation of Quantum Mechanics"', *Proceedings of the Vigier 2000 Symposium*, 21-25 August 2000, Berkeley, CA. [quant-ph/0009070]

Fuchs C.A. 1996. *Distinguishability and Accessible Information in Quantum Theory*, PhD thesis, University of New Mexico, Albuquerque, NM. [quant-ph/9601020].

Fuchs C.A. and Peres A. 1996. 'Quantum-state disturbance versus information gain: Uncertainty relations for quantum information', *Phys. Rev.* A**53**, 2038-2045.

Fuchs C.A. and van de Graaf J. 1999. 'Cryptographic Distinguishability Measures for Quantum Mechanical States', *IEEE Trans. Info. Theory* **45**, 1216 [quant-ph/9712042]

Fuchs C.A. 2001. 'Quantum Foundations in the light of Quantum Information' [quant-ph/0106166], in Gonis 2001 [extended/amended version of this paper: Fuchs 2002]

Fuchs C.A. 2002 'Quantum Mechanics as Quantum Information (and only a little more)', quant-ph/0205039.

Gell-Mann M. and Hartle J.B. 1990. 'Quantum Mechanics in the Light of Quantum Cosmology', in Zurek 1990, 425-458.

Ghirardi G.C., Rimini A. and Weber T. 1980. 'A General Argument against Superluminal Transmission through the Quantum Mechanical Measurement Process', *Lettere al Nuovo Cimento* **27**, 293-8.

Gleason A.M. 1957. 'Measures on the Closed Subspaces of a Hilbert Space', *Journal of Mathematics and Mechanics* **6**, 885-893.

Goldstein H 1980. *Classical Mechanics*, 2nd ed. Addison-Wesley, Reading, MA.

Goldstein S. 1996. 'Bohmian Mechanics and the Quantum Revolution', *Synthese* **107**, 145-165.

Goldstein S.1998. 'Quantum theory without observers', *Physics Today* March: pp.42-46 ; April: pp. 38-42.

Goldwasser S. (ed.) 1994. *Proceedings of the 35th Annual Symposium on the Theory of Computer Science*, IEEE Computer Society Press, Los Alamitos, CA.

Gonis A. (ed.) 2001. *Proceedings of the NATO Advanced Research Workshop on Decoherence and its Implications in Quantum Computation and Information Transfer.*

Greenberger D. *et al.* 1999. *Epistemological and Experimental Perspectives in Quantum Physics.* Kluwer, Dordrecht.

Gromov M. 1987. 'Soft and Hard Symplectic Geometry', in *Proceedings of the International Congress of Mathematicians*, Vol.1-2, AMS, pp. 81-98.

Halvorson H. and Clifton R. 2001. 'Reconsidering Bohr's Reply to EPR', in Butterfield and Placek 2002 [quant-ph 0110107].

Hamming R.W. 1988. *Digital Filters*, 3rd ed.,Prentice Hall, Englewood Cliffs, NJ.

Hardy L. 2001a. 'Quantum Theory From Five Reasonable Axioms' [quant-ph/0101012]

Hardy L. 2001b. ,'Why Quantum Theory?', Contribution to NATO Advanced Research Workshop *Modality, Probability, and Bell's Theorem*, Cracow, Poland, 19.-23.8.01, in Butterfield and Placek 2002 [quant-ph/0111068]

Haroche S. and Raimond J.-M. 1996. 'Quantum Computing: Dream or Nightmare?', Physics Today **49** (8), 51.

Hartle J.B. 1968. 'Quantum Mechanics of Individual Systems, *Am. J. Phys.* **36**, 704-712.

Heelan P.A. 1970a. 'Quantum and Classical Logic: their respective role', *Synthese* **21**, 2-33.

Heelan P.A. 1970b. 'Complementarity, Context Dependence and Quantum Logic', *Found. Phys.* **1**, 95-110 [reprinted in Hooker 1975 Vol.2, pp. 161-180].

Heisenberg W. 1927. 'Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik', *Z. Phys.* **43**, 172-198 – English translation in Wheeler and Zurek 1983, pp. 62-84.

Heisenberg W. 1963. *Physics and Philosophy – The revolution in modern science,* London, Allen and Unwin (1st ed. 1959).

Held C. 1994. 'The Meaning of Complementarity', *Studies in History and Philosophy of Modern Physics* **25**, No. 6, 871-893.

Helstrom C.W. 1976. *Quantum Detection and Estimation Theory*, Mathematics in Science and Engineering, vol. 123, Academic Press, New York.

Hemmo M. and Shenker O. 2003. 'How Maxwell's Demon Defeats Von Neumann', preprint, available at http://philsci-archive.pitt.edu .

Hiley B.J. 1999. 'Active Information and Teleportation', in Greenberger *et al.* 1999, pp.113-126.

Hofer 1998. 'Dynamics, Topology and Holomorphic Curves', in *Proceedings of the International Congress of Mathematicians* Vol.1, DMV, p.255-280.

Holevo A.S. 1973a. 'Information-Theoretical Aspects of Quantum Measurement', *Prob. Info. Trans.* **9**, 110-118.

Holevo A.S. 1973b. 'Bounds for the quantity of information transmitted by a quantum communication channel,' *Prob. Info. Trans.* **9**, 177.

Holland P.R. 1993. *The Quantum Theory of Motion : an account of the de Broglie-Bohm causal interpretation of quantum mechanics*, Cambridge University Press.

Holland P.R. 2001a. 'Hamiltonian Theory of Wave and Particle in Quantum Mechanics', Part I, *Nuovo Cimento* **116B**, 1043-1069.

Holland P.R. 2001b. 'Hamiltonian Theory of Wave and Particle in Quantum Mechanics', Part II, *Nuovo Cimento* **116B**, 1143-1172.

Hooker C.A. 1975 (ed.). *The Logico-Algebraic Approach to Quantum Mechanics.* Vol. 1 & 2, Reidel, Dordrecht.

Horodecki R., Horodecki M. and Horodecki P. 1999. 'Einstein-Podolsky-Rosen Paradox without Entanglement', *Phys. Rev.* A **60**, 4144 (1999) [quant-ph/9811004].

Howson C. and Urbach P. 1993. *Scientific Reasoning: The Bayesian Approach*, 2nd. ed., Open Court, La Salle/Chicago.

Howson C. 2000. *Hume's problem: Induction and the Justification of Belief*, Clarendon Press, Oxford.

Hughes R.I.G. 1989. *The Structure and Interpretation of Quantum Mechanics*, Harvard University Press, Cambridge MA.

Hughston L.P., Jozsa R. and Wootters W.K. 1993. 'A complete classification of quantum ensembles having a given density matrix', *Phys. Lett.* A **183**, 14-18.

Ignatowsky W.V. 1911. *Arch. Math. Phys.* **18**, 17.

Jaeger G. and Shimony A. 1995. *Phys. Let.* A **197**, 83.

Jammer M. 1974. *The Philosophy of Quantum Mechanics. The Interpretation of Quantum Mechanics in Historical Perspective*, John Wiley, New York.

Jammer M. 1989. *The Conceptual Development of Quantum Mechanics*, 2nd ed., Tomash Publishers: American Institute of Physics.

Jauch J.M 1968. *Foundations of Quantum Mechanics*, Addison-Wesley, New York.

Jaynes E.T. 1989. 'Clearing Up Mysteries – The Original Goal', in Skilling (ed.) 1989, pp. 1-27.

Jaynes E.T. 1990. 'Probability in Quantum Theory', in Zurek 1990, pp. 381-403.

Jordan T. 1975. 'Why $-i\nabla$ is the Momentum', *Am. J. Phys.* **43**, 1089-1093.

Jozsa R. 1998. *Quantum Information and its Properties*, in Lo, Popescu and Spiller (eds) 1998, pp. 49-75.

Kadison R. and Ringrose J. 1997. *Fundamentals of the Theory of Operator Algebras*, American Mathematical Society, Providence RI.

Kennedy J.B. 1995. 'On the Empirical Foundations of the Quantum No-Signalling Proofs', *Phil. Sci.* **62**, 543-560.

Khinchin A. 1937. Bull. Univ. Moscow **1**.

Kim Y.-H. , Kulik S.P. and Shih Y. 2001. 'Quantum Teleportation with a Complete Bell State Measurement', *Phys. Rev. Lett.* **86**, 1370[quant-ph/0010046]

Kimura M. 1962. *Diffusion Models in Population Genetics*, Methuen, London.

Kolmogorov A.N. 1933. 'Grundbegriffe der Wahrscheinlichkeitsrechnung', *Ergibnisse der Mathematik* **2**, 196-262.

Koopman B. O. 1931. 'Hamiltonian Systems and Transformations in Hilbert Space', *Proceedings of the National Academy of Sciences of the United States of America* **17**, 315-318.

Kraus, K. 1983. *States, Effects and Operations: Fundamental Notions of Quantum Theory*, Springer-Verlag, Berlin.

Kwiat P.G., Mattle K., Weinfurter H., Zeilinger A., Sergienko A.V. and Shih Y.H. 1995. Phys. Rev. Lett. **75** , 4337.

Kyburg H.E. and Smokler H.E. (eds) 1964. *Studies in Subjective Probability*, Wiley, New York..

Lee A.R. and Kalotas T.M. 1975. 'Lorentz transformations from the first postulate', *Am. J. Phys.* **43**, No.5, 434-437.

Leonhardt U. 1995. 'Quantum-State Tomography and Discrete Wigner Function', *Phys. Rev. Lett.* **74**, 4101.

Lévy-Leblond J.-M. 1974. 'The Pedagogical Role and Epistemological Significance of Group Theory in Quantum mechanics', *Rivista del Nuovo Cimento* **4**, 99-143.

Lévy-Leblond J.-M. 1976. 'One more derivation of the Lorentz transformation', *Am. J. Phys.* **44**, No.3, 271-277.

Liboff R.L. 1984. 'The Correspondence Principle Revisited', *Physics Today* (February), 50-55.

Lo H.K. and Chau H.F. 1997. 'Is Quantum Bit Commitment Really Possible?', *Phys. Rev. Let.* **78**: 3410-3413.

Lo H.K. and Popescu S. 1998. 'The Joy of Entanglement', in Lo, Popescu and Spiller (eds.) 1998, pp. 29-48.

Lo H.K., Popescu S. and Spiller T. (eds) 1998. *Introduction to Quantum Computation and Information*, World Scientific, Singapore.

London F. et Bauer E. 1939. *La théorie de l'observation en mécanique quantique*, Hermann, Paris.

Lukacs E. 1970. *Characteristic Functions*, Charles Griffin, London .

Mackey G.W. 1963. *The Mathematical Foundation of Quantum Mechanics*, Benjamin, New York.

Mann A. and Revzen M. (eds.) 1996. *The Dilemma of Einstein, Podolsky and Rosen - 60 Years Later*, Annals of The Israel Physical Society **12**, 226-257.

Maroney O.J.E. and Hiley B.J. 1999. 'Quantum State Teleportation understood through the Bohm Interpretation', *Found. Phys.* **29**, No.9.

Maroney O.J.E. 2000. 'Sameness and Oppositeness in Quantum Information', ANPA **22**,10-13 August; available at http://www.bbk.ac.uk/tpru/OwenMaroney/paper4/tpru_4.html

Marshall A.W. and Olkin I. 1979. *Inequalities: Theory of majorization and its applications*, Academic Press, New York.

Mattle K., Weinfurter H., Kwiat P.G. , and Zeilinger A. 1996. *Phys. Rev. Lett.* **76** , 4656.

Mayers D. 1997. 'Unconditionally Secure Quantum Bit Commitment is Impossible', *Phys. Rev. Lett.* **78**: 3414-3417.

Mehra J. and Rechenberg H. 1987. *The Historical development of Quantum Theory* (6 vol.), Springer, New York.

Mermin N.D. 1990a. 'Simple Unified Form of the Major No-Hidden-Variables Theorems', *Am. J. Phys.* **65** (27), 3373-3376.

Mermin N.D. 1990b. 'Quantum Mysteries Revisited', *Am.J. Phys.* **58**, 731-734.

Mermin N.D. 1993. *Rev. Mod. Phys.* **65**, 803.

Mermin N.D. 1998. 'What Is Quantum Mechanics Trying to Tell Us?', *Am. J. Phys.* **66**, 753-767 [quant-ph/9801057].

Mermin 1999. 'What Do These Correlations Know about Reality? Nonlocality and the Absurd', *Found. Phys.* **29**, 571-587 [quant-ph/9807055]

Merzbacher E. 1970. Quantum Mechanics (2nd ed.), John Wiley & Sons, New York.

Messiah A. 1961-2. *Quantum Mechanics.* North-Holland, Amsterdam, vol.1-2. [translated from the French by G.M. Temmer]

Miller A.I. (ed.) 1990. *Sixty-Two Years of Uncertainty*, NATO ASI series, Plenum Press, New York.

Miller A.I. 1998. *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905-1911)*, Springer Verlag, New York. (First ed. 1981)

Mittelstaedt P. 1997. *The Interpretation of Quantum Mechanics and the Measurement Process*, Cambridge University Press, New York.

Mott N. 1964. 'On teaching quantum phenomena', *Contemporary physics* **5**, 401-418.

Nelson E. 1985. *Quantum Fluctuations*, Princeton University Press.

von Neumann J. 1932. *Mathematische Grundlagen der Quantenmechanik*, Springer, Berlin [transl. by E.T. Beyer: *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1955]

Nielsen M.A. and Chuang I.L. 2000. *Quantum Information and Quantum Computation*, Cambridge University Press.

Omnès R. 1994. *The Interpretation of Quantum Mechanics*, New Jersey, Princeton University Press.

Parfit D. 1987. 'Divided minds and the nature of persons', in Blakemore and Greenfield 1987.

Penrose R. 1998. 'Quantum computation, entanglement and state reduction', *Phil. Trans. R. Soc. Lond.* A **356**, 1927-1939.

Peres A. 1993. *Quantum Theory: Concepts and methods*, Kluwer, Dordrecht.

Peres A. 1996. *Separability criterion for density matrices*, Phys. Rev. Lett. **77**, 1413.

Peres A. 1999. 'Delayed choice for entanglement swapping', *J. Mod. Optics* (special issue) [quant-ph/9904042]

Peres A. 2002. 'How the No-Cloning Theorem Got Its Name', quant-ph/0205076.

Perina J., Hradil Z. and Jurco B. 1994. *Quantum Optics and Fundamentals of Physics*, Kluwer Academic Publishers, Dordrecht.

Pitowsky I. 1989. *Quantum probability – Quantum logic*, Lecture Notes in Physics 321, Springer Verlag, Berlin.

Popescu S. 1994. 'Bell's Inequality versus Teleportation : What is Nonlocality?', *Phys. Rev. Lett.* **72**, 797-799.

Post E.J. 1995. *Quantum Reprogramming. Ensembles and Single Systems: A Two-Tier Approach to Quantum Mechanics*, Boston Studies in the Philosophy of Science, vol. 181, Kluwer Academic Publishers, Dordrecht.

Przibram K. (ed.) 1967. *Letters on Wave Mechanics*, Philosophical Library, New York.

Redhead M.L.G. 1987. *Incompleteness, Nonlocality and Realism*, Clarendon Press, Oxford.

Robertson H.P. 1929. 'The Uncertainty Principle', *Phys. Rev.* **34**, 163-164.

Rosenfeld L. 1936. 'La première phase de l'évolution de la Théorie des Quanta', *Osiris* **2**, 149-196.

Rovelli C. 1996. 'Relational Quantum Mechanics', *Int. J. Theor. Phys.* **35**, No. 8, 1637-1678 [quant-ph/9609002].

Sanpera A. , Tarrach R. & Vidal G. 1997. 'Quantum separability, time reversal and canonical decompositions', quant-ph/9707041.

Schack R., Brun T.A. and Caves C.M. 2001. 'Quantum Bayes Rule', *Phys.Rev.* A **64**,014305-1-014305-4. [quant-ph/0008113]

Schack R. 2002. 'Quantum Theory from Four of Hardy's Axioms', quant-ph/0210017.

Schilpp P.A. (ed.) 1970. 'Albert Einstein: Philosopher-Scientist, Open Court, La Salle IL, 3rd ed.

Schroeck F. 1996. *Quantum Mechanics on Phase Space*, Kluwer, Dordrecht.

Schrödinger E. 1930. *Proceedings of the Prussian Academy of Sciences*, Physics-Mathematical Section XIX, 296-303 – English translation quant-ph/9903100.

Schrödinger E. 1935a. 'Discussion of Probability Relations Between Separated Systems', *Proc. Cam. Phil. Soc.* **31**, 555-563.

Schrödinger E. 1935b. 'Die gegenwärtige Situation in der Quantenmechanik, *Naturwissenschaften* **23**, No.48, 807-812 No.49, 823-828; No.50, 844-849 (English translation published in *Proc. Am. Phil. Soc.* **124**, 323-338 and reprinted in Wheeler and Zurek 1983, 152-167.

Schrödinger E. 1936. "Probability Relations between Separated Systems", *Proc. Cam. Phil. Soc.* **32**, 446-452.

Schrödinger E. 1951. *Science and Humanism – Physics in our time*, Cambridge University Press.

Schumacher B. 1995. 'Quantum Coding', *Phys. Rev.* A, **51(4)**, 2738-2747.

Scully M.O. and Meystre P. (eds.) 1983. *Quantum Optics, General Relativity, and Measurement Theory*, NATO Advanced Science Institute Series, Plenum Press.

Shannon C.E. 1948. 'A mathematical theory of communication', *Bell Syst. Tech. J.* **27**, 379 & 623. available at:

http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html

Shannon C. E. and Weaver W. 1949. *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

Shenker O. 1999. 'Is -kTr(ρlnρ) The Entropy in Quantum Mechanics?', *The British Journal for the Philosophy of Science* **50**, 33-48.

Shor P. W. 1994. 'Algorithms for quantum computation: discrete logarithms and factoring', in Goldwasser 1994, p. 124.

Skilling J. (ed.) 1989. *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 1989.

Steane A. 1998. 'Quantum computing', *Reports on Progress in Physics* **61**, 117-173 [quant-ph/9708022]

Stinson D.R. 1995. *Cryptography: Theory and Practice*, CRC Press, Boca Raton, Florida.

Stückelberg E.C.G. 1960. 'Quantum Theory in Real Hilbert Space', *Helv. Phys. Acta* **33**, 727-752

Sudbery T. 1997. 'The fastest way from A to B', *Nature* **390**, 551-552.

Synge J.L. and Griffiths B.A. 1959. *Principles of Mechanics* (3rd ed,) McGraw-Hill, New York, London.

Tikochinsky Y. 1988a. 'On the generalized multiplication and addition of complex numbers', *J. Math. Phys.* **29** (2), 398-399.

Tikochinsky Y. 1988b. 'Feynman Rules for Probability Amplitudes', *Int. J. Theor. Phys.* **27** (5), 543-549.

Tribus M. 1961. *Thermostatics and Thermodynamics*, van Nostrand Co., Princeton, NJ.

Uhlmann A. 1976. 'The 'transition probability' in the state space of a *-algebra'. *Reports on Mathematical Physics* **9**, 273-279.

Wehrl A. 1978. 'General Properties of Entropy', *Rev. Mod. Phys.* **50**, 221-259.

Werner R.F. 2001. 'Quantum Information Theory – an Invitation', in Alber *et al.* 2001.

Wheeler J.A. 1982. 'The Computer and the Universe', *Int. J. Theor. Phys.* 21, Nos 6/7

Wheeler J.A. 1983. 'On recognizing "law without law"', *Am. J. Phys.* **51** (5), 398-404.

Wheeler J.A. 1988. 'World as System Self-Synthesized by Quantum Networking', *IBM J. Res. Develop.* **32**, 4-15.

Wheeler J.A.. 1990. 'Information, Physics, Quantum : The Search for Links', in Zurek 1990.

Wheeler J.A and W.H. Zurek (eds.) 1983. *Quantum Theory and Measurement*, Princeton University Press.

Wick G., Wightman A. and Wigner E. 1952. *Phys. Rev.* **88**, 101.

Wigner. E. 1959. *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra*, (translated by J.J. Griffin) Academic Press, New York.

Wigner E.P. 1967. *Symmetries and reflections – Scientific Essays of Eugene P. Wigner*, Indiana University Press, Bloomington.

de Witt B.S. and Graham N. (eds) 1973. *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton University Press.

Wittgenstein L. 1960. *The Blue and Brown Books*, Blackwell, Oxford.

Wootters W.K. 1980. *The Acquisition of Information from Quantum Measurements*, Ph.D. dissertation, University of Texas at Austin.

Wootters W.K. 1981. 'Statistical distance and Hilbert space', *Phys. Rev.* D **23** No.2, 357-362.

Wootters W.K. 1983. 'A measure of the distinguishability of quantum states', in Scully and Meystre 1983.

Wootters W.K. 1986. 'Quantum mechanics without probability amplitudes', *Found. Phys.* **16**(4), 391-405.

Wootters W.K. 1990. 'Local Accessibility of Quantum States', in Zurek (ed.) 1990, pp.39-46.

Wootters W.K. & Zurek W.H. 1982. 'A single quantum cannot be cloned', *Nature* **299**, 802-803.

Zeilinger A. 1997. 'Quantum Teleportation and the non-locality of information', *Phil. Trans. R. Soc. Lond.* A **355**, 2401-2404.

Zeilinger A. 1999. 'A Foundational Principle for Quantum Mechanics', *Found. Phys.* **29**, 631-643.

Zukowski M., Zeilinger A., Horne M.A. and Ekert A.K. 1993. *Phys. Rev. Lett.* **71**, 4287.

Zurek W.H. (ed.) 1990. *Complexity, Entropy and the Physics of Information*, Proceedings of the Santa Fe Institute Studies in the Sciences of Complexity, vol. VIII, Addison-Wesley, Redwood City, CA.