# The Applicability of Shannon Information in Quantum Mechanics and Zeilinger's Foundational Principle*

C.G. Timpson†‡
The Queen's College, Oxford, OX1 4AW, UK

*

†

‡ Thanks are due to Harvey Brown for useful discussion. This work was supported by a studentship from the UK Arts and Humanities Research Board.

**Abstract**

Recently, Brukner and Zeilinger have presented a number of arguments suggesting that the Shannon information is not well defined as a measure of information in quantum mechanics. If established, this result would be highly significant, as the Shannon information is fundamental to the way we think about information not only in classical but also in quantum information theory. On consideration, however, these arguments are found unsuccessful; I go on to suggest how they might be arising as a consequence of Zeilinger's proposed foundational principle for quantum mechanics.

# 1 Introduction

What role the concept of information might have to play in the foundations of quantum mechanics is a question that has recently excited renewed interest (see e.g., Fuchs 2002; Mermin 2002; Wheeler 1990). Zeilinger, for example, has put forward an information-theoretic principle which he suggests might serve as a foundational principle for quantum mechanics (Zeilinger 1999). As a part of this project, Brukner and Zeilinger (2001) have criticised the Shannon measure of information (Shannon 1948), the quantity fundamental to the discussion of information in both classical and quantum information theory. They claim that the Shannon information is not appropriate as a measure of information in the quantum context and have proposed in its stead their own preferred quantity and a notion of 'total information content' associated with it, which latter is supposed to supplant the von Neumann entropy. Their argument takes two forms: first, that the Shannon information is too intimately tied to classical notions of measurement to be applicable in quantum mechanics; and second, that it cannot be used to define an appropriate notion of 'total information content' for quantum systems. I shall argue that neither of these strategies is successful, concentrating rather more on the latter. I shall then try and indicate why these arguments against the Shannon information are arising as a consequence of Zeilinger's proposed foundational principle for quantum mechanics.

## 2 Is the Shannon Information inherently classical?

The Shannon information $H(\vec{p})$ is a measure of uncertainty; it measures the spread of a probability distribution $\vec{p} = \{p_1, \ldots, p_n\}$, quantifying our uncertainty about what the outcome of an experiment described by this distribution will be. It takes the following form:

$$H(\vec{p}) = -\sum_i p_i \log p_i. \tag{1}$$

Brukner and Zeilinger's first concern is that interpreting $H(\vec{p})$ as a measure of information would require a pre-existing sequence of possessed values in a message being decoded, but such a sequence cannot be taken to exist in general in quantum mechanics. (They consider the example of a string of systems all prepared in a given state $|\psi\rangle$ which is not an eigenstate of the observable measured.) Their worry here seems misplaced, however. The possible absence of a pre-existing string of values does not affect the interpretation of the Shannon information, at least as it is usually understood.

Two sorts of explanation standardly relate the Shannon quantity to a notion of information. The first exploits an intuitive link between measures of uncertainty and information: the greater our uncertainty about the outcome of an experiment, the more we stand to gain from actually performing it. From this point of view, the Shannon information tell us our expected information gain before we perform an experiment, or our average gain following many repetitions[1]. Clearly no pre-existing values are called for here, all that is required is that the distribution $\vec{p}$ correctly characterise the experiment in question.

The second link to information follows from Shannon's 1948 noiseless coding theorem, which states that $H(\vec{p})$ tells us the maximum amount that messages drawn from an ensemble characterised by the distribution $\vec{p}$ can be compressed, and hence indicates the channel resources required to transmit messages produced by an information source modelled by such an ensemble. Again there is no requirement that these messages are composed of fixed pre-existing strings of values. To derive the bound on the optimal compression, we consider very long strings of values. Then, given the appropriate probability distribution we know ab initio that any sequence observed will be one of the typical sequences in which the relative frequency of any type of outcome matches its probability of occurrence; and this on its own is sufficient to calculate the number of bits that will be necessary to specify any string produced, pre-existing or not.

Brukner and Zeilinger's second reason for concern that the Shannon information involves problematic classical assumptions is somewhat more substantial. Shannon's original presentation included a uniqueness proof for the form of $H(\vec{p})$ which involved putting forward a number of constraints as reasonable requirements on a measure of uncertainty. The constraint that plays the key role in securing uniqueness is often known as the grouping axiom and takes the following form in the more rigorous presentation of Shannon's argument due to Faddeev (1957):

*For every* $n \geq 2$,

$$H(p_1, p_2, \ldots, p_{n-1}, q_1, q_2) = H(p_1, \ldots, p_{n-1}, p_n) + p_n H(q_1/p_n, q_2/p_n) \qquad (2)$$

*where* $p_n = q_1 + q_2$.

Brukner and Zeilinger argue that if we are to understand the physical content of this axiom then we must refer to the performance of joint experiments. Thus if we take two experiments $A$ and $B$ with outcomes $a_1, \ldots, a_n$; $b_1, \ldots, b_m$ respectively, then the grouping axiom relates our uncertainty for the performance of these distinct experiments. Eqn. (2), they suggest, will be equivalent to

$$H(A \wedge B) = H(A) + H(B|A), \qquad (3)$$

where $H(B|A) = \sum_{i=1}^{n} p(a_i) H(p(b_1|a_i), \ldots, p(b_m|a_i))$.

This, however, seems to make it clear that the grouping axiom embodies a particularly classical assumption about measurement, namely that measurements can be made ideally non-disturbing[2]. Furthermore, it seems that if we are going to be able to apply the grouping axiom, then we have to be able to make the assumption that attributes corresponding to all possible measurements can be assigned to a system simultaneously (in this case, $a_i, b_j$ and $a_i \wedge b_j$). But we know this will not be true in general in quantum mechanics. For non-commuting observables, the probabilities on the left hand sides of eqns.(2) and (3) will not be defined and the grouping axiom will fail to hold. Brukner and Zeilinger thus conclude that the standard uniqueness proof fails in quantum mechanics and that the Shannon information ceases to be justified as a measure of information as it is conceptually tied to these classical assumptions in virtue of the grouping axiom.

The prospects for the Shannon information are not really so bad as this, however. Failure of the argument for uniqueness and inapplicability of the grouping axiom need not imply that the Shannon information cannot function as a measure of uncertainty. The Shannon information is in fact one of a general class of measures of uncertainty, characterised by a set of axioms in which the grouping axiom does not appear (Uffink 1990), hence the grouping axiom is not necessary for the interpretation of the Shannon information as a measure of uncertainty[3] and any classical assumptions that the axiom might embody would not transfer to the Shannon information itself.

It can be argued further that Brukner and Zeilinger's interpretation of the grouping axiom is not equivalent to the standard form which does not involve classical assumptions and is equally applicable in the quantum and classical cases (Timpson 2001). Here, however, I want to note the possibility that Brukner and Zeilinger's worry about the Shannon information, as it finds expression in their grouping axiom argument, may be due at least in part to the fact that they seem to desire a measure of information or uncertainty intrinsic to a quantum system, rather than a measure associated with particular experiments. (This would explain their emphasis on non-commuting measurements on an individual system, for example.) However, if we want a measure of information for a quantum system itself, a measure of how uncertain we are *in general* when we know the state of the system, then it is obvious from the beginning that the Shannon information is not the correct sort of function for us; and there are other, familiar, functions that will do the job instead. A measure of uncertainty is a function of a probability distribution and we know that a joint probability distribution for all possible measurements does not exist; it is for this reason that we introduce measures of mixedness such as the von Neumann entropy which are functions of the *state* rather than of a probability distribution. It would be a mistake, however, to take it as a complaint against the Shannon information that it does not play this very different sort of role. As a measure of uncertainty it does as much as we could ask of it; we do not want to confuse the question of what makes a good measure of uncertainty with the question of when joint probability distributions can be defined.

## 3 The Total Information Constraint

Brukner and Zeilinger's other main argument against the Shannon information is that it is not appropriately related to a notion of 'total information content' for a quantum system. Here they compare $H(\vec{p})$ unfavourably to their preferred quantity

$$I(\vec{p}) = \sum_{i=1}^{n} (p_i - 1/n)^2 \,, \tag{4}$$

which they relate to their notion of total information content in the following way.

A set of measurements is called *mutually unbiased* if the sets of projectors $\{P\}, \{Q\}$ associated with any pair of measurement bases satisfy $\mathrm{Tr}(PQ) = 1/n$, where $n$ is the dimensionality of the system; there can exist at most $n+1$ such bases (Wootters and Fields 1989), constituting a *complete* set[4]. Noting the fact that an unknown state $\rho$ may be completely determined by measurement of such a complete set on an ensemble of similarly prepared systems (Ivanovic 1981), Brukner and Zeilinger suggest that the *total information content* of a quantum system should be defined as a sum of individual information measures for a complete set of mutually unbiased observables. Adopting the measure $I(\vec{p})$, we get:

$$I_{tot} = \sum_{j=1}^{n+1} I(\vec{p^j}) = \sum_{ji} \left(p_i^j - 1/n\right)^2 = \mathrm{Tr}\left(\rho - \mathbf{1}/n\right)^2 . \tag{5}$$

The fact that this quantity is invariant under unitary transformations is important, because Brukner and Zeilinger also suggest that it is a necessary constraint on a measure of total information content that it be unitarily invariant. This is the problem they raise for the Shannon information — substituting $H(\vec{p})$ in eqn. (5) does not result in a unitarily invariant quantity. That is, $H(\vec{p})$ fails to satisfy what we might call Brukner and Zeilinger's 'total information constraint', that a measure of information has to sum to a unitarily invariant quantity that can be interpreted as a 'total information content' for a complete set of mutually unbiased measurements. The picture is that the Shannon measure is inadequate as a measure of information gain because it does not satisfy the total information constraint and hence does not tell us how much of the total information content of a system we learn by performing measurements in a given basis. Similarly, a complaint is raised against the von Neumann entropy that it is merely a measure of mixedness, as unlike $I_{tot}$, it has no relation to the information gained in a measurement unless we happen to measure in the eigenbasis of $\rho$.

This argument against the Shannon information is only compelling if the total information constraint is in fact a reasonable constraint on individual measures of information. Unfortunately, it is not obvious that it is. To begin with, the 'information content' of a quantum state can mean several different things; we might, for example, be interested in the maximum amount that can be encoded into a given quantum system (the Holevo bound (Holevo 1973)), or in the average amount of information needed to specify the state of a system drawn from a given ensemble (the classical information of the source). Since 'information content' is not univocal, it seems difficult to maintain that all meaningful individual information measures and measures of information content must have the particular fixed relation expressed in the total information constraint. A further important consideration is that we may well demand to know why information measures for a complete set of mutually unbiased measurements should be expected to sum to a particularly interesting quantity in any case. To make this question more pointed, let us ask why it is that $I(\vec{p})$ in fact happens to sum to a unitarily invariant quantity.

$I(\vec{p})$ is not especially novel as a measure of information; it is one of the general class of measures of the concentration of a probability distribution given by Uffink (1990). A measure of concentration is the reciprocal of a measure of uncertainty, increasing as a probability distribution becomes more peaked. $I(\vec{p})$ is a Schur convex function, rather than a Schur concave function like $H(\vec{p})$ and measures of uncertainty; it could be said to measure how well we can predict the outcome of an experiment, rather than how uncertain we are about it. $I(\vec{p})$, however, has a particular geometric property as well as being a measure of information; and it is *this* property, tangential to its role as a measure of information, which explains the relation to $I_{tot}$ and how it satisfies the total information constraint. To see this, we will use the Hilbert-Schmidt representation of density operators, a more general form of the familiar Bloch sphere representation.

## 3.1 The Relation between Total Information Content and $I(\vec{p})$

The set of complex $n \times n$ Hermitian matrices forms an $n^2$-dimensional real Hilbert space $V_h(C^n)$ on which we have defined an inner product $(A, B) = \mathrm{Tr}(AB); A, B \in V_h(C^n)$ and a norm $\|A\| = \sqrt{\mathrm{Tr}(A^2)}$ (Fano 1957; Wichmann 1963). The density matrix $\rho$ of an $n$ dimensional quantum system can be represented as a vector in this space. The requirements on $\rho$ of unit trace and positivity imply that the tip of any such vector must lie in the $n^2 - 1$ dimensional hyperplane $\mathbf{T}$ a distance $1/\sqrt{n}$ from the origin and perpendicular to the unit operator $\mathbf{1}$, and on or within a hypersphere of radius one centred on the origin.

It is useful to introduce a set of basis operators on our space; we require $n^2$ linearly independent operators $U_i \in V_h(C^n)$ and it may be useful to require orthogonality: $\mathrm{Tr}(U_i U_j) = \text{const.} \times \delta_{ij}$. Any operator on the system can then be expanded in terms of this basis and in particular, $\rho$ can be written as

$$\rho = \mathbf{1}/n + \sum_{i=1}^{n^2-1} \mathrm{Tr}(\rho U_i) U_i,$$

where we have chosen $U_0 = \mathbf{1}$ to take care of the trace condition.

Evidently, $\rho$ may be determined experimentally by finding the expectation values of the $n^2 - 1$ operators $U_i$ in the state $\rho$. If we include the operator $\mathbf{1}$ in our basis set, then the idempotent projectors associated with measurement of any maximal (non-degenerate) observable will provide a maximum of a further $n - 1$ linearly independent operators. Obtaining the probability distribution for a given maximal observable will thus provide $n - 1$ of the parameters required to determine the state, and the minimum number of measurements of maximal observables that will be needed in total is $n + 1$, if each observable provides a full complement of linearly independent projectors.

Each such set of projectors spans an $n - 1$ dimensional hyperplane in $V_h(C^n)$ and their expectation values specify the projection of the state $\rho$ into this hyperplane. Ivanovic (1981) noted that projectors $P, Q$ belonging to any two different mutually unbiased bases will be orthogonal in $\mathbf{T}$, hence the hyperplanes associated with measurement of mutually unbiased observables are orthogonal in the space in which density operators are constrained to lie in virtue of the trace condition. If $n + 1$ mutually unbiased observables can be found, then, $V_h(C^n)$ can be decomposed into orthogonal subspaces given by the one dimensional subspace spanned by $\mathbf{1}$ and the $n + 1$ subspaces associated with the mutually unbiased observables. The state $\rho$ can then be expressed as:

$$\rho = \mathbf{1}/n + \sum_{j=1}^{n+1} \sum_{i=1}^{n} q_i^j \bar{P}_i^j, \tag{6}$$

where $\bar{P}_i^j = P_i^j - \mathbf{1}/n$ is the projection onto $\mathbf{T}$ of the $i$th idempotent projector in the $j$th mutually unbiased basis set, and $q_i^j = (p_i^j - 1/n)$ is the expectation value of this operator in the state $\rho$. For a given value of $j$, the vectors $\bar{P}_i$ span an $(n-1)$ dimensional orthogonal subspace and the square of the length of a vector expressed in the form (6) lying in subspace $j$ will be given by $\sum_{i=1}^{n}(q_i^j)^2 = I(\vec{p^j})$.

It is then simple to see that $I(\vec{p})$ satisfies the total information constraint because these squared lengths of the components of $\rho$ in orthogonal spaces can just be added to get the length squared of $\rho$ in $\mathbf{T}$, i.e. the square of the distance of $\rho$ from the maximally mixed state $\mathrm{Tr}(\rho - \mathbf{1}/n)^2 = I_{tot}$; and this is what eqn. (5) reports.

Thus $I(\vec{p})$ satisfies the total information constraint because it has the particular geometrical property of measuring a length. The question now is, would $H(\vec{p})$ have to be a measure of length in order to be a measure of information? That is, does it suffer from not satisfying the total information constraint? The short answer is no — $H(\vec{p})$ can be a perfectly good measure of information without having to be a measure of the length of the projection of $\rho$ into the subspace associated with an observable. The longer answer involves pointing out that when considered strictly as measures of information, $I(\vec{p})$ and $H(\vec{p})$ function in much the same way; and in fact, as measures of information, $H(\vec{p})$ stands to $S(\rho)$, the von Neumann entropy, in the same relation as $I(\vec{p})$ stands to $I_{tot}$, as we shall now see.

Brukner and Zeilinger's total information content $I_{tot}$ seems best interpreted as a measure of mixedness, a measure of how much we know in general about what the outcomes of experiments will be given the state (Brukner and Zeilinger 1999b). The functioning of measures of mixedness can usefully be approached via the notions of majorization and Schur convexity (concavity). The majorization relation $\prec$ imposes a pre-order on probability distributions (Uffink 1990; Nielsen 2001). A probability distribution $\vec{q}$ is majorized by $\vec{p}$, $\vec{q} \prec \vec{p}$, iff $q_i = \sum_j S_{ij} p_j$, where $S_{ij}$ is a doubly stochastic matrix. That is (via Birkhoff's theorem), if $\vec{q}$ is a mixture of permutations of $\vec{p}$. Thus if $\vec{q} \prec \vec{p}$, then $\vec{q}$ is a more mixed or disordered distribution than $\vec{p}$.

Schur convex (concave) functions respect the ordering of the majorization relation: a function $f$ is Schur convex if, if $\vec{q} \prec \vec{p}$ then $f(\vec{q}) \leq f(\vec{p})$, and Schur concave if, if $\vec{q} \prec \vec{p}$ then $f(\vec{q}) \geq f(\vec{p})$ (for strictly Schur convex(cave) functions, equality holds only if $\vec{q}$ and $\vec{p}$ are permutations of one another). This explains the utility of such functions as measures of the concentration and uncertainty of probability distributions, respectively. Now, it can be shown (Nielsen 2001) that the probability distribution $\vec{p}$ for the outcomes of any projective measurement is majorized by the vector of eigenvalues $\vec{\lambda}$ of the pre-measurement state $\rho$. This entails that $S(\rho)$ is the infimum of $H(\vec{p})$, $H(\vec{p}) \geq S(\rho)$ (since $H(\vec{p})$ is Schur concave), and $I_{tot}$ is the supremum of $I(\vec{p})$, $I(\vec{p}) \leq I_{tot}$ ($I(\vec{p})$ Schur convex); both inequalities reflecting the same fact from the theory of majorization. These relations illustrate why a measure of mixedness is a measure of how much we know given the state:

the more mixed a state, the more uncertain we must be about the outcome of any given measurement. However, they also make it clear that $S(\rho)$ does have an explicit relation to the information gain from measurement that would justify its interpretation as a total information content. Conversely, they establish that $H(\vec{p})$ does have an appropriate relation to a measure of information content, despite not satisfying the total information constraint; the same relation, in fact, that $I(\vec{p})$ has to *its* associated notion of information content, up to an irrelevant change in sign. We must conclude that the total information constraint is not a reasonable requirement on measures of information; the Shannon information survives Brukner and Zeilinger's final argument unscathed.

# 4 Zeilinger's Foundational Principle

We have seen that Brukner and Zeilinger's worries about the applicability of the Shannon information are misplaced; the Shannon information is perfectly well defined and meaningful as a measure of information in quantum mechanics. I want now to suggest that these worries may have arisen in the first place as a consequence of a proposed foundational principle for quantum mechanics.

Zeilinger (1999) puts forward the following principle as a possible foundation for the whole of quantum theory. Two formulations of the Principle are presented:

**FP1**) *An elementary system represents the truth value of one proposition.*

**FP2**) *An elementary system carries one bit of information.*

Brukner and Zeilinger claim that this Principle can explain, amongst other things, the irreducible randomness of quantum measurement and the phenomenon of entanglement.

It is not immediately obvious that FP1 and FP2 are actually equivalent, neither is it clear how the Foundational Principle might in fact be supposed to function. As it stands it does not appear to distinguish between classical and quantum; FP1 and FP2 seem to be as true of a single classical (Ising model) spin as of a qubit. Unfortunately, space does not allow us to discuss properly the prospects for the Foundational Principle as a foundational principle here (see Timpson (2001)), we shall have to rest content with trying to become a little clearer on what it actually means. To this end, we need to discuss Zeilinger's conception of the quantum state and to elaborate what he means by a system 'carrying', or 'representing' information.

Zeilinger adopts an explicitly instrumentalist view of the quantum state:

> The initial state...represents all our information as obtained by earlier observation...[the time evolved] state is just a short hand way of representing the outcomes of all possible future observations.(Zeilinger 1999, 634)

Such instrumentalist sentiments are common. Where Brukner and Zeilinger depart from the norm, however, is in adopting a very literal construal of the information taken to constitute the state, by adopting, at least inchoately, the Hilbert-Schmidt representation of states:

> We describe a photon by a catalog of information ("information vector") $\vec{i} = (i_1, i_2)$ about mutually complementary propositions $\{\mathcal{P}_1, \mathcal{P}_2\}$. Such propositions are, for example, $\mathcal{P}_1$: "the polarization of the photon is vertical (horizontal)" (Brukner and Zeilinger 1999a)

The component $i_1$ is defined as $(p - q)$, where $p$ and $q$ are the probabilities for vertical and horizontal polarization respectively[5]. Thus, the components of the information vector $\vec{i}$ correspond, effectively, to the coefficients $q_i^j$ in eqn. (6), and the propositions $\mathcal{P}$ to the operators $\bar{P}_i^j$.

On this conception, an amount of information in the form of probabilities has been associated to propositions representing the outcomes of mutually unbiased measurements; the information and the experimental propositions it is about can be read off directly from the Hilbert-Schmidt representation of the state, given some choice of basis operators (choice of complete set of mutually unbiased measurements). Illustrating the general idea, if probability 1 is associated to some proposition, then the state says the maximum possible about the outcome of the measurement with which that proposition is associated; if there is a flat distribution for outcomes of a measurement, the state contains no information about it. In general the state will contain *partial* information about a number of mutually unbiased observables. Endorsing the instrumentalist line, *all* that the state is is an amount of information in this way about mutually complementary observables.

Now the statements FP1 and FP2 refer to an elementary system carrying or representing an amount of information. By this, Zeilinger says, he means the following:

> ...that a system "represents" the truth value of a proposition or that it "carries" one bit of information only implies a statement concerning what can be said about possible measurement results. (Zeilinger 1999, 635)

Thus rather than, for example, being a restriction on how much information might be encoded into, or read from, a physical system, we see that the Foundational Principle is a restriction on how much can be said about measurement outcomes, and hence, in particular, is a restriction on how much the *state* can say about measurement outcomes. For Zeilinger, the state will in general be constituted by amounts of partial information about measurement outcomes. The Foundational Principle requires that the state can only contain a limited amount of information, namely one bit; hence it follows that the amounts of partial information contained in the state, although how these are to be quantified has not yet been specified in detail, must add up to one bit's worth in total.

This, however, rules out the Shannon information as the measure of the amount 'carried' by the state about a given measurement; we know that in general we will not have a sum to unity for amounts of partial information conceived in the way outlined. (As $H(\vec{p})$ does not sum to a unitarily invariant quantity for a complete set of mutually unbiased measurements, we cannot guarantee that we will attain the value of one for any given pure state.)

Thus the conjunction of the Foundational Principle with Brukner and Zeilinger's brand of literal instrumentalism about the quantum state is inconsistent with adopting the Shannon information to measure the amount of information 'carried' about a measurement. I suggest that it is this fact that tempts Brukner and Zeilinger to argue, unsuccessfully as it turns out, that the Shannon information is not the correct measure of information and cannot be applied in quantum mechanics.

We close with two final comments. First, consider what someone rather more realist about the quantum state might make of the Foundational Principle. Here the information idiom would no longer be particularly enticing and a more precise statement of what is being expressed by the Foundational Principle would be natural:

> **'R' FP)** *Any projective measurement other than in the eigenbasis of $\rho$ results in a shorter vector in $V_h(C^n)$*

('R' FP for 'realist' Foundational Principle.) That is, any such measurement would result in a more spread probability distribution; if we began with a pure state then post- (non-selective) measurement, the ensemble will no longer be represented by a one-dimensional projector. Given this statement of the Principle, we see that it is a matter of choice whether or not, or with which quantities, we chose to discuss the uncertainties associated with the probability distributions generated by the state.

Second, we might wonder whether the foregoing indicates that for the instrumentalist at least, $I(\vec{p})$ does after all represent the 'correct' measure of information in quantum mechanics. Such a choice would appear very artificial given the close relation between the functioning of $I(\vec{p})$ and $H(\vec{p})$ discussed earlier. Note, however, that one could still be an instrumentalist about the quantum state while adopting 'R'FP as more genuinely informative than FP1 and FP2. The instrumentalist is not, then, forced to accept $I(\vec{p})$ as the only correct measure of information in quantum mechanics.

So, to conclude: we have seen that Brukner and Zeilinger's arguments against the applicability of the Shannon information in quantum mechanics are unsuccessful; and we have seen, moreover, that these arguments seem to be motivated by the conjunction of Zeilinger's Foundational Principle with a particular form of instrumentalism about the quantum state. Even if one has instrumentalist leanings, however, this does not imply that the Brukner-Zeilinger measure can be the only correct measure of information in quantum mechanics. The Shannon information remains perfectly well defined and meaningful as a measure of information in the quantum context.

# References

Brukner, Caslav, and Anton Zeilinger (1999a), "Malus' Law and Quantum Information", *Acta Physica Slovaka* 49 (4): 647–52.

— (1999b) "Operationally Invariant Information in Quantum Measurements", *Physical Review Letters* 83 (17): 3354.

— (2001), "Conceptual Inadequacy of the Shannon Information in Quantum Measurements", *Physical Review A* 63: 022113.

Faddeev, D. K (1957), in H Grell (ed.) *Arbeiten zum Informationstheorie I*. Berlin: Deutscher Verlag der Wissenschaften, 88–91.

Fano, Ugo (1957), "Description of States in Quantum Mechanics by Density Operator Techniques", *Reviews of Modern Physics* 29 (1): 74–93.

Fuchs, Christopher A (2002), "Quantum Mechanics as Quantum Information (and only a little more)" `arXiv:quant-ph/0205039`.

Holevo, Alexander S. (1973), "Information Theoretical Aspects of Quantum Measurement", *Problems of Information Transmission (USSR)* 9: 177.

Ivanovic, I. D. (1981), "Geometrical Description of Quantal State Determination", *Journal of Physics A* 14: 3241–3245.

Jaynes, Edwin T (1957), "Information Theory and Statistical Mechanics", *Physical Review* 106 (4): 620–30.

Mermin, N. David (2002), “Whose Knowledge?”, in Reinhold Bertlmann and Anton Zeilinger (eds.), “Quantum (Un)speakables: Essays in Commemoration of John S. Bell”. Berlin, Heidleberg: Springer-Verlag, `arXiv:quant-ph/0107151`.

Nielsen, Michael A. (2001), “Characterizing Mixing and Measurement in Quantum Mechanics”, *Physical Review A* 63: 022114.

Shannon, Claude E. (1948), “The Mathematical Theory of Communication”, *Bell Systems Technical Journal* 27: 379–423, 623–656. Reprinted in Claude E. Shannon and Warren Weaver (eds.), *The Mathematical Theory of Communication*. Urbana and Chicago: University of Illinois Press, 30-125.

Timpson, Christopher G. (2001), “On a Supposed Conceptual Inadequacy of the Shannon Information in Quantum Mechanics”, To appear in *Studies in the History and Philosophy of Modern Physics*, `arXiv:quant-ph/0112178`.

Uffink, Jos (1990), *Measures of Uncertainty and the Uncertainty Principle*. Ph.D. Dissertation, Utrecht, Netherlands: University of Utrecht.

Wheeler, John A (1990), “Information, Physics, Quantum: The Search for Links”, in Wojciech Zurek (ed.) “Complexity, Entropy and the Physics of Information”. Redwood City, CA: Addison-Wesley, 3–28.

Wichmann, Eyvind H. (1963), “Density Matrices Arising from Incomplete Measurements”, *Journal of Mathematical Physics* 4 (7): 884–896.

Wootters, William K., and Brian D. Fields (1989), "Optimal State-Determination by Mutually Unbiased Measurements", *Annals of Physics (NY)* 191: 363.

Zeilinger, Anton (1999), "A Foundational Principle for Quantum Mechanics", *Foundations of Physics* 29 (4): 631–43.

## Notes

[1] $H(\vec{p})$ actually takes the form of an expectation value of a function, $-\log p_i$, that decreases the more likely an outcome is.

[2] Eqn. (3) can be read as saying that if we first perform $A$ and then $B$, our uncertainty in $B$ can just be updated conditional on the $A$ outcome, our ability to predict $B$ values not being degraded by the $A$ measurement.

[3] Uffink (1990, §1.6.3) argues further that the grouping axiom is not a natural constraint on a measure of information, even in the classical case, and should not be insisted upon as a necessary constraint, *pace* Jaynes (1957).

[4] For the $n = 2$ case, the three spatially orthogonal components of spin constitute a familiar example of a complete set of mutually unbiased observables.

[5] For this two-dimensional quantum system, we have here, essentially, the Bloch sphere representation.