Conditional Action and Quantum Versions of Maxwell's Demon

Heinz-Jürgen Schmidt[1]

Abstract

We propose a new way of looking at the quantum Maxwell's demon problem in terms of conditional action. A "conditional action" on a system is a unitary time evolution, selected according to the result of a previous measurement, which can reduce the entropy of the system. However, any conditional action can be realized by an (unconditional) unitary time evolution of a larger system and a subsequent Lüders measurement, whereby the entropy of the entire system is either increased or remains constant. We give some examples that illustrate and confirm our proposal, including the erasure of N qubits and the Szilard engine, thus relating the present approach to the Szilard principle and the Landauer principle that have been discussed as possible solutions of the Maxwell's demon problem.

Keywords Maxwell's demon · Quantum measurement

1 Introduction

Since its first appearance in 1867, the thought experiment of James Clerk Maxwell has given rise to many ideas and probably more than a 1000 papers [1–3][1]. In the thought experiment a demon controls a small door between two gas chambers. When single gas molecules approach the door, the demon opens and closes the door quickly, so that only fast molecules enter one of the chambers, while slow molecules enter the other one. In this way the demon's behavior causes one chamber to heat up and the other to cool down, reducing entropy and violating the 2nd law of thermodynamics.

Among the most influential defenses of the 2nd law are those of Szilard [4] and Landauer and Bennett [5, 6]. Szilard proposes his own version ("Szilard's engine")

---

[1] For an overview of work on Maxwell's demon see [1–3]

✉ Heinz-Jürgen Schmidt
hschmidt@uos.de

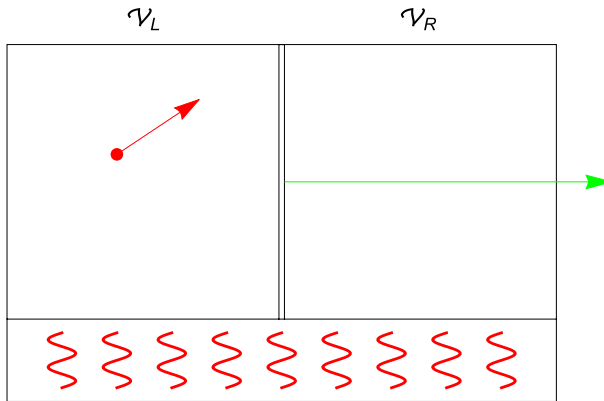[1] Universität Osnabrück, Fachbereich Physik, 49069 Osnabrück, Germany

Fig. 1 Schematic representation of Szilard's engine. A volume is separated by a piston into two chambers $\mathcal{V}_R$ and $\mathcal{V}_L$ of equal volume. A molecule is localized by a measurement in one of these chambers and, depending on the result of the measurement, an isothermal expansion of the one-molecule gas in contact with a heat bath will be performed to the right or to the left. In the figure we show the case where the molecule has been found in the left chamber and the expansion is performed to the right

of the original thought experiment that consists only of one gas particle which can be found in the right or the left chamber of a cylindrical box divided by a piston. Depending on its position an isothermal expansion of the one-molecule gas is performed to the left or to the right thereby converting heat from a heat bath completely into work, see Fig. 1. Szilard argues that the entropy decrease of the system is compensated by the entropy costs of acquiring information about the position of the gas particle ("Szilard's principle"). His arguments are formulated within classical physics and not easy to understand, see also the analysis and reconstruction of Szilard's reasoning in [3, 7].

Based on Landauer's calculations [5] on the thermodynamics of computing Bennett has shifted the focus from the entropy costs of acquiring to *erasing* information [6]. He argues that for a cyclic operation of a Szilard engine converting heat completely into work the memory device that contains the information about the initial measurement should be set to a default value each time. This erasure of information produces at least the entropy needed to compensate the entropy decrease caused by the engine. This explanation ("Landauer's principle") has today been adopted by the main stream of physicists, but has also been criticized by a minority of scholars, see [2, 3, 8] and further references cited there. For the present paper it will be sensible to distinguish between the principle that erasure of memory produces entropy ("Landauer's principle" in the narrow sense) and the position that this effect constitutes the solution of the apparent paradox of Maxwell's demon (henceforward called "Landauer/Bennett principle").

Whereas the arguments of Szilard and Landauer/Bennett are mainly classical, it appears plausible that a proper account of entropy increase due to measurements should be discussed within the realm of quantum theory. A first attempt of a quantum-theoretical account of Szilard's engine has been given by Zurek [9], followed by [10–13]. More recently, the paradigm of Maxwell's demon has been

used in connection with quantum information theory, especially quantum error correction, see [14] and [15].

Zurek in his [9] considered a one-particle quantum system in a box described by a Gibbs ensemble and calculated the increase of free energy due to the measurement of whether the particle is in the right or in the left chamber. In the section of his paper headlined "Measurement by 'quantum Maxwell's demon' " Zurek presented a model of the measurement using ideas of decoherence and finally also incorporated the Landauer/Bennett issue of memory erasure. However, the complete entropy balance remains opaque. In terms of content, it would be plausible to regard the paper as a quantum mechanical justification of the Szilard principle. But then the statement in the summary

> "Moreover, we show that the ultimate reason for the entropy increase can be traced back to the necessity to 'reset' the state of the measuring apparatus, which, in turn, must involve a measurement."

would appear as an unfounded tribute to the Landauer/Bennett principle. Therefore the general message is not quite clear. Further, there are three questions left open:

- Are the information-theoretic concepts used in [9] only an illustration of the theoretical account or are they crucial to solve the Maxwell's demon problem? This question is the more important since there exist suggestions of extending the framework of statistical mechanics by information-theoretic notions, see, e. g., [16, 17].
- Similarly, are the ideas from theories of decoherence, see also [10] and [11], really necessary to solve the Maxwell's demon problem?
- Since the paper follows very closely the details of Szilard's engine, one wonders which assumptions and approximations are decisive for the solution presented and which are only made for convenience. In other words, a more abstract representation of the "quantum Maxwell's demon" would be desirable.

In the present paper we will pursue a similar approach but try to amend and extend Zurek's results in the way indicated above. Our explanation of the apparent paradoxical results of Maxwell's demon acting on a quantum system (also called "object system") will be given in three steps:

- First we define the concept of "conditional action" that comprises the original version of Maxwell's demon as well as Szilard's engine and Landauer's erasure of memory. The mathematical representation of "conditional action" on quantum systems results in a special kind of *instruments*, in the sense of [20], that we will call "Maxwell instruments".
- We show that the total operation of a Maxwell instrument *may* decrease the von Neumann entropy of the object system depending on the initial state. If this happens we will call the Maxwell instrument "demonic".
- A demonic Maxwell instrument always has a physical realization of the following kind: The object system is extended by an auxiliary system and the

total system undergoes a unitary time evolution followed by a Lüders meas-
urement at the auxiliary system. If reduced to the object system the final state
will have a smaller entropy than at the beginning although the total entropy
will increase in accordance with what a 2nd law of quantum mechanics pre-
sumably would predict.

It has been criticized [2, 3] that the Landauer/Bennett defense of the 2nd law against
Maxwell's demon in turn presupposes the 2nd law. We avoid these pitfalls of circu-
larity since we do not assume any general 2nd law in quantum mechanics but only
a few well-established theorems about the increase of von Neumann entropy dur-
ing Lüders measurements and state separation. Actually we would not know how
to formulate such a general 2nd quantum law. In this respect the role of Maxwell's
thought experiment is different in classical and in quantum theory: In classical the-
ory it is a potential paradox since it seems to contradict the well-established 2nd
law. In quantum theory it is rather a tool to find such a general 2nd law. Fortunately,
Maxwell-demonic interventions can be formalized within the realm of quantum
measurement theory where already fragments of a 2nd law exist that are sufficient to
explain the demon's actions.

The paper is organized as follows: In Sect. 2 we recapitulate some well-known
definitions and results from quantum measurement theory for the convenience of the
reader. These concepts are applied in Sect. 3 to explain why the conditional action
of Maxwell' demon possibly lowers the entropy of the object system but leads to an
at least equal amount of entropy increase in some auxiliary system. The following
Sect. 4 contains two simple examples illustrating the former considerations. A clas-
sical version of "conditional action" will be sketched in Sect. 5, followed by a Sum-
mary in Sect. 6. We have deferred some proofs (A, B) and the explicit construction
of a measurement dilation of a Maxwell instrument (C) into the Appendix, as well
as the detailed account of Szilard's engine (D) according to our approach.

2 Operations and Instruments

In the following sections we will heavily rely upon the mathematical notions of
operations and *instruments*. Although these notions are well-known it will be in
order to recall the pertinent definitions adapted to the present purposes and their
interpretations in the context of measurement theory. In order to keep the presenta-
tion as simple as possible we restrict ourselves to the case of finite dimensional Hil-
bert spaces $\mathcal{H}$ and refer the reader to the literature on the general case of separable
Hilbert spaces.

Let $B(\mathcal{H})$ denote the space of Hermitean operators $A : \mathcal{H} \longrightarrow \mathcal{H}$ and $B_+(\mathcal{H})$ the
cone of positively semi-definite operators, i. e., having only non-negatives eigenval-
ues. Moreover, let $T : B(\mathcal{H}) \longrightarrow B(\mathcal{H})$ be a linear map satisfying

- T is *positive*, i. e., maps $B_+(\mathcal{H})$ into itself,
- T is *completely positive*. This means that $T \otimes \mathbb{1} : B(\mathcal{H} \otimes \mathbb{C}^N) \longrightarrow B(\mathcal{H} \otimes \mathbb{C}^N)$
  will be positive for all positive integers N.

Then T will be called an *operation*. It may be trace-preserving or not.

Operations are intended to describe state changes due to measurements. By definition, a *Lüders measurement* (without selection according to the outcomes) induces the state change

$$\rho \mapsto L(\rho) = \sum_{n \in \mathcal{N}} P_n \, \rho \, P_n \, , \tag{1}$$

where $(P_n)_{n \in \mathcal{N}}$ denotes a complete family of mutually orthogonal projections $P_n \in B_+(\mathcal{H})$. The Lüders operation L is an example of a trace-preserving operation. Note that the map (1) is defined for all $\rho \in B(\mathcal{H})$ whereas the physical interpretation holds only for *statistical operators* ρ, i. e., for positively semi-definite operators with $\mathrm{Tr}(\rho) = 1$.

We mention the following representation theorem for operations, see, e. g., [20, Proposition 7.7], or [15, Chapter 8.2.3]. A is an operation iff it can be written as

$$A(\rho) = \sum_{i \in \mathcal{I}} A_i \, \rho \, A_i^* \, , \tag{2}$$

with the *Kraus operators* $A_i : \mathcal{H} \to \mathcal{H}$ and a finite index set $\mathcal{I}$. Comparison of (1) and (2) shows that for the Lüders operation one may choose $\mathcal{I} = \mathcal{N}$ and $A_n = P_n$ for all $n \in \mathcal{N}$.

In (1) we have considered the total state change without any selection. If we select according to the outcome of the Lüders measurement we would obtain a family of (not trace preserving) operations

$$L_n(\rho) = P_n \, \rho \, P_n, \quad n \in \mathcal{N}, \tag{3}$$

that describe conditional state changes. This situation can be generalized in the following way.

Let $\mathcal{N}$ be a finite index set. Then the map $\mathfrak{I} : \mathcal{N} \times B(\mathcal{H}) \longrightarrow B(\mathcal{H})$ will be called an *instrument* iff

- $\mathfrak{I}(n)$ is an operation for all $n \in \mathcal{N}$, and
- $\mathrm{Tr}\left(\sum_{n \in \mathcal{N}} \mathfrak{I}(n)(\rho)\right) = \mathrm{Tr}\rho$ for all $\rho \in B(\mathcal{H})$.

The second condition can be rephrased by saying that the *total operation* $\mathfrak{I}(\mathcal{N})$ defined by

$$\mathfrak{I}(\mathcal{N})(\rho) \equiv \sum_{n \in \mathcal{N}} \mathfrak{I}(n)(\rho) \tag{4}$$

will be trace-preserving. The special case (3) will be referred to as a *Lüders instrument*.

The comparison with the definition 7.5 of [20] shows that, besides neglecting convergence conditions, we have specialized the general definition of an instrument to the case of a finite outcome space $\mathcal{N}$. Measurements of continuous observables like position or momentum would require to consider elements of the σ-algebra of

Borel subsets of, say, $\mathbb{R}^N$ for the first argument of the instrument. This generalization is not necessary to be considered in the present paper.

We will need a second representation theorem, this time formulated for instruments. It is called a *measurement dilation* and can be physically viewed as a realization of a non-Lüders instrument $\mathfrak{J}$ by a time evolution and a Lüders instrument on a larger system. Thus let $\mathcal{K}$ be another Hilbert space, $\phi \in \mathcal{K}$ a vector with $\|\phi\| = 1$ and corresponding projection P_ϕ and $V : \mathcal{H} \otimes \mathcal{K} \longrightarrow \mathcal{H} \otimes \mathcal{K}$ a unitary operator. Further, let $(Q_n)_{n \in \mathcal{N}}$ be a complete family of mutually orthogonal projections in $\mathcal{K}$. Then the map $\mathfrak{D}_{\mathcal{K}, \phi, V, Q} : \mathcal{N} \times B(\mathcal{H}) \longrightarrow B(\mathcal{H})$ defined by

$$\mathfrak{D}_{\mathcal{K}, \phi, V, Q}(n)(\rho) \tag{5}$$

$$\equiv \mathrm{Tr}_{\mathcal{K}}\big((\mathbb{1} \otimes Q_n) V (\rho \otimes P_\phi) V^* (\mathbb{1} \otimes Q_n) \big) \tag{6}$$

will be an instrument. Here $\mathrm{Tr}_{\mathcal{K}}$ denotes the *partial trace* that reduces a state of the total system to a state of the subsystem given by the Hilbert space $\mathcal{H}$. If $\mathfrak{J}$ is a given instrument then $\mathfrak{D}_{\mathcal{K}, \phi, V, Q}$ will be called a *measurement dilation* of $\mathfrak{J}$ iff $\mathfrak{J} = \mathfrak{D}_{\mathcal{K}, \phi, V, Q}$. The mentioned representation theorem guarantees the existence of measurement dilations for any given instrument, see Theorem 7. 14 of [20] or Exercise 8. 9 of [15]. The last reference also contains an explicit construction procedure for $\mathfrak{D}_{\mathcal{K}, \phi, V, Q}$ that will be reproduced for the special case of a Maxwell instrument in Appendix C and will henceforward be referred to as the "standard realization".

3 The Quantum Version of Maxwell's Demon (QMD)

The activity of Maxwell's demon can be abstractly characterized as performing a *conditional action*, i. e., an action depending on the results of a previous measurement. Additionally, it is required that this conditional action leads to an entropy decrease of the system if applied to a certain set $\mathcal{A}$ of *admissible initial states*. In this paper we will interpret these notions quantum mechanically, especially the states as statistical operators ρ of a so-called object system defined on some Hilbert space $\mathcal{H}$, and the measurement as a Lüders instrument

$$\mathfrak{J}(n)(\rho) = P_n \rho P_n \,, \tag{7}$$

where n runs through some finite index set $\mathcal{N}$ and $(P_n)_{n \in \mathcal{N}}$ is a complete family of mutually orthogonal projections. The total Lüders operation

$$\mathfrak{J}(\mathcal{N})(\rho) = \sum_{n \in \mathcal{N}} P_n \rho P_n \tag{8}$$

represents the state change after the Lüders measurement without any selection. More general instruments may be used to model the demon's measurement but this possibility will not be considered in the present paper.

Further, the entropy is taken as the von Neumann entropy [18]

$$S(\rho) = -\mathrm{T}r(\rho \log \rho) \,, \tag{9}$$

where log is chosen as the natural logarithm. It is well-known [15, 18, 23] that the entropy of a state never decreases during a Lüders measurement, i. e.,

$$S(\rho) \le S\left(\sum_{n \in \mathcal{N}} P_n \, \rho \, P_n \right) \equiv \widetilde{S}_1 \,. \tag{10}$$

Hence a Lüders measurement alone cannot be used to model a QMD. Additionally, we need to give a quantum-theoretical definition of a *conditional action* relative to a Lüders measurement. This will be done by considering a family $\left(U_n \right)_{n \in \mathcal{N}}$ of unitary operators in $\mathcal{H}$ such that the combined state change will be given by the instrument

$$\mathfrak{F}(n)(\rho) = U_n P_n \, \rho P_n \, U_n^* \,, \tag{11}$$

henceforward called a "Maxwell instrument", with total operation ("Maxwell operation")

$$\rho \mapsto \rho_1 = \mathfrak{F}(\mathcal{N})(\rho) = \sum_{n \in \mathcal{N}} U_n P_n \, \rho P_n \, U_n^* \,. \tag{12}$$

Again the Kraus operators $A_n = U_n P_n$ of the operation $\mathfrak{F}(\mathcal{N})$ may be read off the representation (12).

We stress that we will use the mathematical notion of an instrument that was originally designed to characterize state changes due to measurements in order to describe the more general state changes caused by a measurement *and* a conditional action. A similar approach has been adopted in chapter 12.4.4 of [15] in connection with quantum error correction.

It can be shown that a Maxwell operation always decreases the entropy of the corresponding post-measurement state:

Proposition 1

$$S_1 \equiv S\left(\sum_{n \in \mathcal{N}} U_n P_n \, \rho P_n \, U_n^* \right) \le S\left(\sum_{n \in \mathcal{N}} P_n \, \rho P_n \right) = \widetilde{S}_1 \,. \tag{13}$$

For a proof see Appendix B.

It is obvious that the U_n are not uniquely determined by (11), for example, U_n must only be defined on the support of P_n and can be arbitrarily extended to its orthogonal complement. In other words: the conditional action must be only defined for those cases where the condition holds.

In passing we note that the concept of "conditional action" is also used in quantum teleportation, see [15], chapter 1.3.7. Here Alice makes two quantum measurements and sends her results to Bob via a classical communication channel, who in turn performs certain operations depending on the measurement results. However, the total entropy increases during teleportation and hence it cannot be considered as a QMD.

It is well-known that in the case of a more general instrument than that of Lüders type a statement analogous to (10) may fail, i. e., a generalized measurement can decrease entropy, see [15], Exercise 11.15. We will provide two examples in Sect. 4 showing that this may also happen for an instrument of the form (12) and hence the Maxwell instrument is a possible candidate for a QMD.

We know from classical thermodynamics that the decrease of entropy of some system would not contradict the 2nd law of thermodynamics if it is accompanied by an equal or larger increase of entropy in some other parts of the world. This strategy of explaining the decrease of entropy can also be tried in the case of quantum mechanics. It is highly plausible that the demon needs some auxiliary system to perform the measurement and the conditional action. We will call this auxiliary system again the "demon" and assume that it can be modelled as another quantum system with Hilbert space $\mathcal{K}$. How can the quantum demon be realized? It is tempting to use the *measurement dilation* sketched in Sect. 2 that was originally intended to merely give a physical realization of a non-Lüders measurement. But there is no reason not to apply this construction to Maxwell instruments $\mathfrak{J}$ as well.

Hence we will assume that at the beginning the state of the combined system, object system and demon, is assumed to be

$$\rho \otimes P_\phi \, , \tag{14}$$

where P_ϕ is a one-dimensional projector in $\mathcal{K}$. Then a unitary time evolution V of the combined system takes place with the resulting state being

$$V \left(\rho \otimes P_\phi \right) V^* \, , \tag{15}$$

followed by a Lüders measurement at the demon with projectors $Q_n : \mathcal{K} \to \mathcal{K}$. This leads to a (not normalized) state

$$\left(\mathbb{1} \otimes Q_n \right) V \left(\rho \otimes P_\phi \right) V^* \left(\mathbb{1} \otimes Q_n \right) \, . \tag{16}$$

Finally this state is reduced to the object system by performing the partial trace $\mathrm{Tr}_{\mathcal{K}}$. This yields the measurement dilation of $\mathfrak{J}$ of the form

$$\begin{aligned} &\mathfrak{D}_{\mathcal{K},\phi,V,Q}(n)(\rho) \\ &\equiv \mathrm{Tr}_{\mathcal{K}} \left(\left(\mathbb{1} \otimes Q_n \right) V \left(\rho \otimes P_\phi \right) V^* \left(\mathbb{1} \otimes Q_n \right) \right) \, , \end{aligned} \tag{17}$$

with corresponding total operation

$$\begin{aligned} &\mathfrak{D}_{\mathcal{K},\phi,V,Q}(\mathcal{N})(\rho) \\ &= \sum_{n \in \mathcal{N}} \mathrm{Tr}_{\mathcal{K}} \left(\left(\mathbb{1} \otimes Q_n \right) V \left(\rho \otimes P_\phi \right) V^* \left(\mathbb{1} \otimes Q_n \right) \right) \end{aligned} \tag{18}$$

Before entering into the proposed solution of the mentioned paradox we would like to point out that the measurement dilation (17) in a sense reverses the temporal order of measurement and (conditional) action. In the original description of the demon we imagine a measurement followed by an action depending on the result of that measurement. In the dilation (17) there is first an unconditioned time evolution of

the combined system followed by a state change due to a Lüders measurement at the demon and the state reduction. This resembles the difference between a classical computer that executes an "if-else" command thereby performing a conditional action and a quantum computer that performs all possible actions simultaneously until a final measurement selects which condition is satisfied. Such a realization seems strange at first sight but is a consequence of our decision to describe the demon purely as a quantum system.

Coming back to the apparent violation of a tentative 2nd law it is clear that the entropy of the quantum state remains constant during the first steps of the operation $\mathfrak{D}(\mathcal{N})$:

$$S_0 \equiv S(\rho) = S\left(\rho \otimes P_\phi\right) = S\left(V\left(\rho \otimes P_\phi\right) V^*\right), \tag{19}$$

since the entropy is additive for tensor products, vanishes for pure states and is unitarily invariant. By the following Lüders measurement the entropy increases (or remains constant) according to (10):

$$S(\rho) \leq S(\rho_{12}), \tag{20}$$

$$\rho_{12} \equiv \sum_{n \in \mathcal{N}} \left(\mathbb{1} \otimes Q_n\right) V \left(\rho \otimes P_\phi\right) V^* \left(\mathbb{1} \otimes Q_n\right). \tag{21}$$

If we reduce ρ_{12} to both subsystems,

$$\rho_{12} \mapsto \rho_1 \otimes \rho_2 \equiv \left(\mathrm{Tr}_{\mathcal{K}}\rho_{12}\right) \otimes \left(\mathrm{Tr}_{\mathcal{H}}\rho_{12}\right), \tag{22}$$

the entropy further increases:

$$S_0 \leq S(\rho_{12}) \leq S(\rho_1) + S(\rho_2). \tag{23}$$

This is a consequence of the so-called subadditivity of the von Neumann entropy, see [15], 11.3.4. The inequality (23) is compatible with the condition for a QMD

$$S(\rho_1) < S_0, \tag{24}$$

since it only implies

$$S(\rho_2) \overset{(23)}{\geq} S_0 - S(\rho_1) \overset{(24)}{>} 0. \tag{25}$$

This means that the decrease of the entropy of the object system will be, at least, compensated by an increase of the demon's entropy. In this case the total entropy of the object system and the demon does not decrease in accordance with a tentative 2nd law.

4 Examples

4.1 Erasure of *N* Qubits

As a first example of a demonic Maxwell instrument $\mathfrak{E}$ and its standard realization we consider a system with a Hilbert space being an *N*-fold tensor product of two-dimensional ones

$$\mathcal{H} = \bigotimes_{\nu=1}^{N} \mathbb{C}^2_{(\nu)} \tag{26}$$

and an orthonormal basis of vectors $|n\rangle$, $n \in \mathcal{N} \equiv \{0, \dots, 2^N - 1\}$ where n is identified with the string of length N consisting of its binary digits. Especially, 0 represents the string consisting of N zeroes. Further we choose an initial Lüders measurement with projectors

$$P_n = |n\rangle\langle n|, \ n \in \mathcal{N}, \tag{27}$$

and the unitaries U_n corresponding to the Maxwell instrument (11) such that

$$U_n |n\rangle = |0\rangle \tag{28}$$

for all $n \in \mathcal{N}$. After a short calculation we obtain

$$\mathfrak{E}(\mathcal{N})(\rho) = \sum_{n\in\mathcal{N}} U_n P_n \rho P_n U_n = P_0, \tag{29}$$

for all statistical operators ρ in $\mathcal{H}$ and hence the description of the Maxwell instrument $\mathfrak{E}$ as "erasure of N qubits" seems adequate. Since

$$S(\mathfrak{E}(\mathcal{N})(\rho)) = S(P_0) = 0, \tag{30}$$

the entropy decrease of the corresponding Maxwell operation is maximal and we may call it "demonic".

Its standard realization is given by $\mathcal{K} = \mathcal{H}$, $\phi = |0\rangle$, $Q_n = P_n$ for all $n \in \mathcal{N}$ and

$$V : \mathcal{H} \otimes \mathcal{K} \longrightarrow \mathcal{H} \otimes \mathcal{K}, \quad V(\phi_1 \otimes \phi_2) = \phi_2 \otimes \phi_1. \tag{31}$$

After a short calculation we obtain, in accordance with (29),

$$\rho_1 = \mathrm{Tr}_{\mathcal{K}}(\rho_{12}) = P_0, \tag{32}$$

where

$$\rho_{12} \equiv \sum_{n\in\mathcal{N}} (\mathbb{1} \otimes Q_n) V (\rho \otimes P_0) V^* (\mathbb{1} \otimes Q_n), \tag{33}$$

and

$$\rho_2 = \mathrm{Tr}_{\mathcal{H}}(\rho_{12}) = \sum_{n \in \mathcal{N}} P_n \rho P_n . \tag{34}$$

Moreover,

$$S(\rho_2) = S\left(\sum_{n \in \mathcal{N}} P_n \rho P_n \right) \geq S(\rho) , \tag{35}$$

by virtue of (10).

This means that the standard realization of the Maxwell instrument $\mathfrak{E}$ erasing N qubits proceeds by shifting the post-measurement state of the Lüders measurement corresponding to (27) into an auxiliary system of the same size as the object system. According to (35) this overcompensates the decrease of entropy due to the erasure. Since we have not precisely stated a quantum version of Landauer's principle (in the narrow sense) we cannot claim that this would represent a proof of this principle. A possible obstacle would be that such a principle is usually formulated to make a statement about *all* possible realizations of the erasure process, whereas we have only said what would be obtained for realizations by measurement dilations $\mathfrak{E} = \mathfrak{D}_{\mathcal{K}, \phi, V, Q}$.

Note finally that the usual statement about the entropic costs for erasure of at least $k_{\mathrm{B}} \log 2$ per bit (re-introducing the Boltzmann constant k_{B}) follows from

$$S\left(\sum_{n \in \mathcal{N}} P_n \rho P_n \right) = - \sum_{n \in \mathcal{N}} p_n \log p_n , \tag{36}$$

if all $p_n \equiv \mathrm{Tr}(\rho P_n)$ are equal and hence $p_n = 2^{-N}$ which entails $-\sum_{n \in \mathcal{N}} p_n \log p_n = N \log 2$.

4.2 A Simple Model of a QMD

Similarly as in the case of Szilard's engine [4] we simplify the QMD scenario to a one-particle problem. Further, we consider only two pairs of yes-no-properties of the particle:

- Position: right or left (r/l),
- Speed: hot or cold (h/c).

This leads to a 4-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^4$ spanned by the four orthogonal states $|rh\rangle, |rc\rangle, |lh\rangle, |lc\rangle$. For the Lüders measurement we assume

$$P_1 = |rh\rangle\langle rh|, \quad P_2 = \mathbb{1} - P_1, \quad \mathcal{N} = \{1, 2\}. \tag{37}$$

As the conditional action we choose

$$U_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad U_2 = \mathbb{1} . \tag{38}$$

This means that, if the particle is found at the right hand side and being hot then it is transferred to the left hand side without changing its speed:

$$U_1|rh\rangle = |lh\rangle . \tag{39}$$

The action of U_1 onto the other three basis vectors is irrelevant since it models a conditional action and will only be applied in the case where the first Lüders measurement has the result "yes" and yields the post measurement state $|rh\rangle$. If the measurement result is "no" then U_2 will be applied, i. e., there will be no action.

Next we restrict the class $\mathcal{A}$ of admissible initial states to those of the form

$$\rho = \frac{p}{2}(|rh\rangle\langle rh| + |lh\rangle\langle lh|) + \frac{1-p}{2}(|rc\rangle\langle rc| + |lc\rangle\langle lc|) , \tag{40}$$

where $0 < p < 1$. This means that initially the particle is in a mixed state with probability p of being "hot" irrespective of its position. It follows that initially the entropy will be

$$S_0 = S(\rho) = -\left(p \log \frac{p}{2} + (1-p)\log \frac{1-p}{2} \right) . \tag{41}$$

The final state ρ_1 according to (12) will be

$$\rho_1 = p\,|rh\rangle\langle rh| + \frac{1-p}{2}(|rc\rangle\langle rc| + |lc\rangle\langle lc|) \tag{42}$$

having the entropy

$$S_1 = S(\rho_1) = -\left(p \log p + (1-p)\log \frac{1-p}{2} \right) . \tag{43}$$

Comparison with (41) yields

$$S_1 - S_0 = -p \log 2 < 0 , \tag{44}$$

and hence the model is a proper QMD since the action of the demon leads to a decrease of the object system's entropy.

Our next aim is to construct a measurement dilation of the form (17) following the prescription given in Appendix C. Hence we choose $\mathcal{K} = \mathbb{C}^2$ with standard basis $\phi = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\psi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, and

$$Q_1 = P_\phi, \quad Q_2 = P_\psi . \tag{45}$$

The linear operators in $\mathcal{H} \otimes \mathcal{K} = \mathbb{C}^4 \otimes \mathbb{C}^2 \cong \mathbb{C}^4 \oplus \mathbb{C}^4$ will be represented by 2×2-matrices the entries of which are 4×4-matrices. This simplifies the calculation of partial traces. With this convention we set

$$
V = \left(\begin{array}{cccc|cccc}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}\right). \tag{46}
$$

One may confirm by direct calculation that with the above definitions $\mathfrak{D}_{\mathcal{K},\phi,V,Q}$ is indeed a measurement dilation of the considered Maxwell instrument.

Additionally, we will explicitly calculate the measurement dilation for admissible initial states stepwise using the fact that all states will be diagonal in the standard basis of $\mathbb{C}^4 \oplus \mathbb{C}^4$. First we note that

$$
\rho \otimes P_\phi = \operatorname{diag}\left(\frac{p}{2}, \frac{1-p}{2}, \frac{p}{2}, \frac{1-p}{2}, 0, 0, 0, 0\right). \tag{47}
$$

Since $V(\rho \otimes P_\phi)V^*$ is already diagonal we obtain

$$
\rho_{12} = V(\rho \otimes P_\phi)V^* \tag{48}
$$

$$
= \sum_{n=1}^{2} (\mathbb{1} \otimes Q_n) \, V(\rho \otimes P_\phi)V^* \, (\mathbb{1} \otimes Q_n) \tag{49}
$$

$$
= \operatorname{diag}\left(0, 0, \frac{p}{2}, 0, 0, \frac{1-p}{2}, \frac{p}{2}, \frac{1-p}{2}\right). \tag{50}
$$

From this we obtain the partial trace ρ_1 as the sum of the two diagonal blocks of ρ_{12}:

$$
\rho_1 = \operatorname{Tr}_{\mathcal{K}} \rho_{12} = \operatorname{diag}\left(0, \frac{1-p}{2}, p, \frac{1-p}{2}\right) \tag{51}
$$

in accordance with (42) and its entropy (43). Analogously, the final state of the demon is obtained by taking the traces of the block matrices and has the form

$$
\rho_2 = \operatorname{Tr}_{\mathcal{H}} \rho_{12} = \operatorname{diag}\left(\frac{p}{2}, 1 - \frac{p}{2}\right) \tag{52}
$$

with entropy

$$
S_2 = S(\rho_2) = -\left(\frac{p}{2} \log \frac{p}{2} + \left(1 - \frac{p}{2}\right) \log \left(1 - \frac{p}{2}\right)\right). \tag{53}
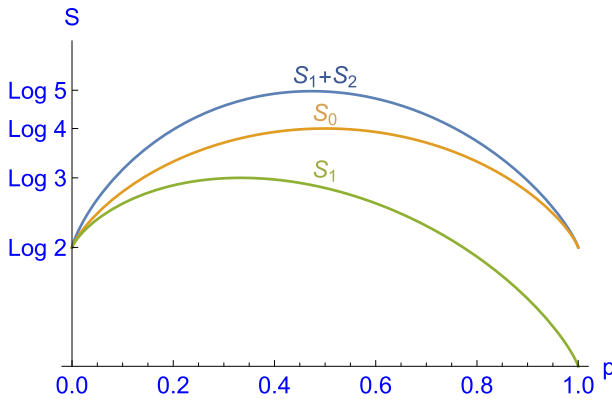$$

Fig. 2 Plot of various entropies of the simple QMD model as functions of the probability p: The initial entropy S_0 of the total system ($=$ the initial entropy of the object system) (orange curve), the final entropy $S_1 + S_2$ of the total system (blue curve), and the final entropy S_1 of the object system (green curve). We see that always $S_1 < S_0$ but $S_1 + S_2 > S_0$ if $0 < p < 1$

This leads to

$$S_1 < S_0, \quad \text{but } S_1 + S_2 > S_0 , \tag{54}$$

see Fig. 2, and hence the decrease of entropy of the object system is overcompensated by the increase of the demon's entropy in our example.

A remarkable detail of our example is the fact that the state of the combined system after the interaction

$$V\left(\rho \otimes P_\phi\right)V^* \tag{55}$$

commutes with all projections $\mathbb{1} \otimes Q_n$ and hence the entropy increase due to the Lüders measurement vanishes. The final entropy increase is completely due to the separation of the total state into reduced states of the subsystems. It has been argued against Szilard's principle that there are also reversible measurements and hence this principle alone is not sufficient to defense the 2nd law against the Maxwell's demon objection, see [6], chapter 5. Our example yields a counter argument closely related to Zurek's consideration of mutual information [9]: In the quantum case there are also entropy costs of state separation that might suffice to compensate the entropy decrease of the object system even if the measurement is reversible (adiabatic).

5 Classical Conditional Action

It will be instructive to investigate the classical counterpart of the conditional action relative to a (Lüders) measurement introduced in Sect. 3. To this end we consider probability distributions

$$p : \mathcal{I} \to [0, 1] \tag{56}$$

defined on a finite set $\mathcal{I}$ of elementary events and subject to the condition

$$\sum_{i \in \mathcal{I}} p_i = 1 \, . \tag{57}$$

A "measurement" will be represented by a partition of $\mathcal{I}$, i. e., a disjoint union

$$\mathcal{I} = \biguplus_{j \in \mathcal{J}} I_j \, . \tag{58}$$

As usual, we define the Shannon entropy [19], up to a factor $\log 2$, by

$$H(p) \equiv - \sum_{i \in \mathcal{I}} p_i \log p_i \, . \tag{59}$$

Then a "classical conditional action" relative to the measurement $\left(I_j \right)_{j \in \mathcal{J}}$ will be defined by a map

$$\phi : \mathcal{I} \to \mathcal{I} \, , \tag{60}$$

that is injective on the subsets I_j, i. e., if $i_1, i_2 \in I_j$ for some $j \in \mathcal{J}$ and $i_1 \neq i_2$ then $\phi(i_1) \neq \phi(i_2)$ holds. Each conditional action gives rise to a new probability distribution $q : \mathcal{I} \to [0, 1]$ defined by

$$q_i \equiv \sum_{\phi(k)=i} p_k \, , \tag{61}$$

that has, in contrast to the quantum case, *always* a lower (or the same) entropy:

$$H(q) \leq H(p) \, . \tag{62}$$

Proof of Eq. 62 If ϕ is a global bijection then (62) is satisfied with equality. Now assume that exactly two events are mapped onto the same one, say, $\phi(1) = \phi(2) = i$ and $p_1, p_2 > 0$. Then we conclude, for $j = 1, 2$,

$$\log p_j < \log(p_1 + p_2), \tag{63}$$

$$- \log p_j > - \log(p_1 + p_2), \tag{64}$$

$$-p_1 \log p_1 - p_2 \log p_2 > - (p_1 + p_2) \log(p_1 + p_2) \tag{65}$$

$$= - q_i \log q_i \, , \tag{66}$$

which means that the fusion of two probabilities p_1 and p_2 to q_i decreases the corresponding term of the entropy. From this the general case follows by induction. ☐

We will give an elementary example. Let $\mathcal{I} = \{1, 2, 3, 4, 5, 6\}$ denote the numbers of a die and $p_i = 1/6$ their probabilities. The measurement detects whether the dice

roll is low or high, corresponding to the partition $\mathcal{I} = I_1 \uplus I_2 = \{1, 2, 3\} \uplus \{4, 5, 6\}$. If the dice roll is low, the die is flipped so that the new roll is high. If the dice roll is already high, nothing is done. This describes the conditional action

$$\phi(i) = \begin{cases} i & \text{if } i \in I_2, \\ 7 - i & \text{if } i \in I_1. \end{cases} \tag{67}$$

The new probability distribution q generated by the conditional action will be given by $q_1 = q_2 = q_3 = 0$ and $q_4 = q_5 = q_6 = 1/3$. It has the entropy $H(q) = \log 3 < H(p) = \log 6$, in accordance with (62).

Returning to the general case we will define the analogue of the "measurement dilation" considered in Sect. 3. The first step is to consider the extended event space

$$\Omega = \mathcal{I} \times \mathcal{J} \tag{68}$$

and a fixed initial value $j_0 \in \mathcal{J}$. This means that the initial distribution $P_2 : \mathcal{J} \to [0, 1]$ is concentrated on the value j_0 and hence has vanishing entropy, $H(P_2) = 0$.

Define the injective map $\Phi : \mathcal{I} \times \{j_0\} \to \Omega$ by

$$\Phi(i, j_0) \equiv (\phi(i), \mathrm{j}(i)) , \tag{69}$$

where we have written $j = \mathrm{j}(i)$ if $i \in I_j$. The injectivity of Φ follows since $i_1 \neq i_2$ and $i_1, i_2 \in I_j$ for some $j \in \mathcal{J}$ implies $\phi(i_1, j_0) \neq \phi(i_2, j_0)$ by the assumption that ϕ is injective on I_j. If i_1, i_2 lie in different sets I_j then $\mathrm{j}(i_1) \neq \mathrm{j}(i_2)$). Hence Φ can be extended to a bijective map $\bar{\Phi} : \Omega \to \Omega$ that is the analogue of the unitary operator V introduced in Eq. (15).

Φ maps p onto a new probability distribution Q on Ω defined by

$$Q(k, j) = \begin{cases} p_i & : \text{if } \Phi(i, j_0) = (k, j), \\ 0 & : \text{else,} \end{cases} \tag{70}$$

with the same entropy, $H(Q) = H(p)$. Let Q_1 denote the first marginal distribution of Q given by

$$Q_1(k) = \sum_{j \in \mathcal{J}} Q(k, j) , \tag{71}$$

and, analogously,

$$Q_2(j) = \sum_{k \in \mathcal{I}} Q(k, j) . \tag{72}$$

Then it can be shown that Q_1 coincides with the distribution q defined above. The proof uses

$$Q_1(k) \overset{(71)}{=} \sum_{j \in \mathcal{J}} Q(k, j) \tag{73}$$

$$\overset{(70)}{=} \sum_{j \in \mathcal{J}, \Phi(i,j_0)=(k,j)} p_i \tag{74}$$

$$\overset{(69)}{=} \sum_{j \in \mathcal{J}, (\phi(i),j(i))=(k,j)} p_i \tag{75}$$

$$= \sum_{\phi(i)=k} p_i \tag{76}$$

$$\overset{(61)}{=} q(k) . \tag{77}$$

By the subadditivity of the Shannon entropy, see [15] Theorem 11.3 (4), we have $H(Q_1) + H(Q_2) \geq H(Q)$ and hence $H(Q_2) \geq H(Q) - H(Q_1) = H(p) - H(q)$. This means that the entropy decrease $H(q) - H(p) < 0$ due to the conditional action is (over)compensated by entropy increase of $H(Q_2) - H(P_2) = H(Q_2)$, analogously to the quantum case.

In order to illustrate the measurement dilation for the above example, we first note that $\mathcal{J} = \{1, 2\}$ can be viewed as a kind of memory of whether the die has been flipped ($j = 2$) or not ($j = 1$). Let $j_0 \equiv 1$, then the map Φ is given by

$$\begin{aligned} \Phi(1, 1) &= (6, 2), \ \Phi(2, 1) = (5, 2), \ \Phi(3, 1) = (4, 2), \\ \Phi(4, 1) &= (4, 1), \ \Phi(5, 1) = (5, 1), \ \Phi(6, 1) = (6, 1). \end{aligned} \tag{78}$$

The resulting probability distribution Q satisfies

$$\begin{aligned} Q(4, 1) &= Q(5, 1) = Q(6, 1) \\ &= Q(4, 2) = Q(5, 2) = Q(6, 2) = 1/6 , \end{aligned} \tag{79}$$

and vanishes for other events. Hence $H(Q) = \log 6$. The marginal distributions are obtained as $Q_1 = q$ and $Q_2(1) = Q_2(2) = 1/2$. Hence $H(Q_1) = \log 3$ and $H(Q_2) = \log 2$. The latter exactly compensates the entropy decrease $H(q) - H(p) = -\log 2$ due to the conditional action.

6 Summary

We have given an explanation of the apparently paradoxical entropy decrease of a quantum system caused by the external intervention analogous to but more general than Maxwell's demon. This explanation follows Szilard's principle [4] and its quantum version given by Zurek [9] in so far as it includes the demon's state into the entropy balance. But we extend these approaches by introducing the concept of "conditional action" and its mathematical description in terms of a "Maxwell instrument". The quantum-mechanical description of the demon can then be accomplished by using tools from quantum measurement theory [20], especially the "measurement dilation" of a Maxwell instrument. The entropy decrease due to the conditional

action of Maxwell's demon thus appears as a special case of the entropy decrease due to a non-Lüders measurement and has an analogous explanation, see [21, 22] or [23]. Of course, we have not shown that *all* physical realizations of Maxwell's demon would be compatible with a tentative 2nd law, but only those described by measurement dilations.

The relation of our explanation to the Landauer/Bennett principle proves to be ambivalent. On the one hand there is no contradiction: If the conditional action is intended to be part of a cyclic process it would be necessary to reset the state of the demon to its initial value. This is only possible by another conditional action performed by a second demon and ends up with an increased entropy of the second demon's state. But on the other hand it would not be entirely appropriate to call this process an "erasure of memory" since in our approach the function of the demon cannot be reduced to a mere memory, but also includes the role of a measuring device and of a control unit for the conditional action. Moreover, the reset of the demon's state was motivated by getting started a cyclic process. If this reset necessarily increases the entropy of some other part of the environment, this simply means that it has not achieved its goal and hence is superfluous. From this perspective the Landauer/Bennett principle appears as a possible supplement to Szilard's principle but can hardly be viewed as "the ultimate reason for the entropy increase" [9].

It has been argued [3] that current explanations of Maxwell's demon using principles connecting information and entropy are not yet based on firm grounds. It is therefore worth mentioning that our approach does not rely on concepts from information theory, notwithstanding the frequent citation of a textbook [15] on quantum information theory and the use of von Neumann entropy. One may object, what is information anyway, if not the result of measurements used to trigger conditional action? But what one is actually concerned with here is the methodological distinction between specialization and generalization. It may be possible to introduce new concepts that fit specific situations without extending the theory in question. However, this must be strictly separated from the situation where new terms and laws are required to generalize the theory. Conceptual parsimony can be helpful to clearly distinguish between these two cases.

Appendix A: Characterization of Maxwell Instruments

There exists a so-called statistical duality between states and observables, see [20], chapter 23.1. In the finite-dimensional case $B(\mathcal{H})$ can be identified with its dual space $B(\mathcal{H})^*$ by means of the Euclidean scalar product $\mathrm{Tr}\,(A\,B)$. Physically, we may distinguish between the two spaces in the sense that $B(\mathcal{H})$ is spanned by the subset of statistical operators representing *states* and $B(\mathcal{H})^*$ is spanned by the subset of operators with eigenvalues in the interval $[0, 1]$ representing *effects*. Effects describe yes-no-measurements including the subset of projectors, which are the extremal points of the convex set of effects, see [20].

Every operation $A : B(\mathcal{H}) \rightarrow B(\mathcal{H})$, viewed as a transformation of states (Schrödinger picture) gives rise to the dual operation $A^* : B(\mathcal{H})^* \longrightarrow B(\mathcal{H})^*$

viewed as a transformation of effects (Heisenberg picture). Reconsider the representation (2) of the operation A by means of the Kraus operators A_i. Then the dual operation A^* has the corresponding representation

$$A^*(X) = \sum_{i \in \mathcal{I}} A_i^* X A_i \,, \tag{A1}$$

for all $X \in B(\mathcal{H})^*$.

Similarly, every instrument $\mathfrak{I}$ gives rise to a dual instrument $\mathfrak{I}^* : \mathcal{N} \times B(\mathcal{H})^* \longrightarrow B(\mathcal{H})^*$ defined by

$$\mathfrak{I}^*(n)(X) \equiv \mathfrak{I}(n)^*(X) \tag{A2}$$

for all $n \in \mathcal{N}$ and $X \in B(\mathcal{H})^*$. The condition that the total operation (4) will be trace-preserving translates into

$$\mathfrak{I}^*(\mathcal{N})(\mathbb{1}) = \sum_{n \in \mathcal{N}} \mathfrak{I}^*(n)(\mathbb{1}) = \mathbb{1} \,. \tag{A3}$$

Thus every dual instrument yields a resolution of the identity by means of effects $F_n = \mathfrak{I}^*(n)(\mathbb{1})$, and hence to a generalized observable in the sense of a *positive operator-valued measure* (POM) [20]. The traditional notion of "sharp" observables represented by self-adjoint operators corresponds to the special case of a projection-valued measure P_n, $n \in \mathcal{N}$, satisfying $\sum_{n \in \mathcal{N}} P_n = \mathbb{1}$.

An operation $A : B(\mathcal{H}) \to B(\mathcal{H})$ will be called *pure* iff it maps rank one operators onto rank one operators. Physically, this means that a pure operation maps pure states onto pure states, up to a positive factor. If the representation (2) of A can be reduced to a single Kraus operator, i. e.,

$$A(\rho) = A_1 \rho A_1^* \,, \tag{A4}$$

then A will be a pure operation. Conversely, every pure operation has a representation of the form (A4), as can be shown by means of lemma 7.8 in [24] (note that this reference does not require complete positivity for operations). *Pure instruments* are defined analogously. Note that Maxwell instruments will be pure since they consist of pure operations according to (11).

Then we can formulate the following characterization of Maxwell instruments:

Proposition 2 *An instrument $\mathfrak{I}$ will be a Maxwell instrument iff it is pure and gives rise to a sharp observable, i. e., $P_n \equiv \mathfrak{I}^*(n)(\mathbb{1})$ will be a projection for all $n \in \mathcal{N}$.*

Proof "only-if-part": As remarked above, a Maxwell instrument $\mathfrak{I}$ will be pure and, according to (11), its dual has the representation

$$\mathfrak{I}^*(n)(X) = P_n U^* X U P_n \,, \tag{A5}$$

for all $n \in \mathcal{N}$ and $X \in B(\mathcal{H})^*$. Hence

$$\mathfrak{I}^*(n)(\mathbb{1}) = P_n U^* \mathbb{1} U P_n = P_n \,, \tag{A6}$$

will be a projector for all $n \in \mathcal{N}$.

"if-part": If $\mathfrak{I}(n)$ is pure it has the representation (A4) and hence

$$\mathfrak{I}^*(n)(X) = A_n^* X A_n , \tag{A7}$$

for all $n \in \mathcal{N}$ and $X \in B(\mathcal{H})^*$. Let $A_n^* = Q_n U_n^*$ be the polar decomposition of A_n^* such that Q_n is positively semi-definite and U_n^* unitary. It follows by assumption that

$$\mathfrak{I}^*(n)(\mathbb{1}) = A_n^* A_n = \left(Q_n U_n^*\right)\left(U_n Q_n\right) = (Q_n)^2 \tag{A8}$$

is a projector for all $n \in \mathcal{N}$ and hence we may set $Q_n = P_n$ and obtain

$$\mathfrak{I}(n)(\rho) = A_n \rho A_n^* = U_n P_n \rho P_n U_n^* , \tag{A9}$$

for all $n \in \mathcal{N}$ and $\rho \in B(\mathcal{H})$. This shows that $\mathfrak{I}$ is a Maxwell instrument and completes the proof of Proposition 2. $\qquad\square$

Appendix B: Conditional Action Decreases Entropy

Proof of Proposition 1 Define

$$p_n \equiv \mathrm{Tr}\left(\rho P_n\right) , \tag{B1}$$

$$\rho_n \equiv \frac{1}{p_n} P_n \rho P_n , \tag{B2}$$

and

$$\widetilde{\rho}_n \equiv \frac{1}{p_n} U_n P_n \rho P_n U_n^* = U_n \rho_n U_n^* , \tag{B3}$$

for all $n \in \mathcal{N}$. Obviously,

$$S(\widetilde{\rho}_n) = S(\rho_n) . \tag{B4}$$

Since the ρ_n have orthogonal support, theorem 11.8 (4) of [15] can be applied and yields:

$$\widetilde{S}_1 = S\left(\sum_{n \in \mathcal{N}} p_n \rho_n\right) = \sum_{n \in \mathcal{N}} p_n S(\rho_n) + H(p) , \tag{B5}$$

where $H(p)$ is the Shannon entropy, see (59). Analogously, theorem 11.10 of [15] yields

$$S_1 = S\left(\sum_{n \in \mathcal{N}} p_n \widetilde{\rho}_n\right) \leq \sum_{n \in \mathcal{N}} p_n S(\widetilde{\rho}_n) + H(p) \tag{B6}$$

$$\overset{(B4)}{=} \sum_{n \in \mathcal{N}} p_n S(\rho_n) + H(p) \tag{B7}$$

$$\overset{(B5)}{=} S\left(\sum_{n \in \mathcal{N}} p_n \rho_n\right) = \widetilde{S}_1 \,, \tag{B8}$$

which completes the proof of Proposition 1.   □

If the initial state ρ and the family of projections P_n, $n \in \mathcal{N}$, is given, one may ask which choice of the unitary operators U_n, $n \in \mathcal{N}$, would minimize the entropy $S_1 = S\left(\sum_{n \in \mathcal{N}} U_n P_n \rho P_n U_n^*\right)$? We conjecture the following result.

Let $\left(\psi_\mu\right)_{\mu \in \mathcal{M}}$ be an orthonormal basis in $\mathcal{H}$ and $\left(\phi_\mu^{(n)}\right)_{\mu=1,\dots,d_n}$ an eigenbasis of $P_n \rho P_n$ such that

$$P_n \rho P_n \phi_\mu^{(n)} = r_\mu^{(n)} \phi_\mu^{(n)} \tag{B9}$$

for all $\mu = 1, \dots, d_n$ and $n \in \mathcal{N}$. We assume that the order of the indices μ is chosen such that the eigenvalues of $P_n \rho P_n$ are monotonically decreasing:

$$r_1^{(n)} \geq r_2^{(n)} \geq \dots \geq r_{d_n}^{(n)} \tag{B10}$$

for all $n \in \mathcal{N}$. Then an optimal choice of the U_n is given by the conditions

$$U_n \phi_\mu^{(n)} = \psi_\mu \tag{B11}$$

for all $\mu = 1, \dots, d_n$ and $n \in \mathcal{N}$. This means that the U_n merge the eigenspaces of $P_n \rho P_n$ as much as possible such that the largest corresponding eigenvalues are added thereby decreasing the entropy of the state. The above choice is not unique since, e. g., global permutations of the eigenvalues do not change the entropy.

Of course, it is not clear in general whether this decrease of entropy leads to $S_1 < S_0$. Only in the latter case we would call the resulting Maxwell instrument "demonic". If the choice of the P_n, $n \in \mathcal{N}$, also remains open the problem becomes trivial: Upon choosing the P_n, $n \in \mathcal{N}$, one-dimensional the above optimal choice of the U_n yields a pure state with vanishing entropy, as in the case of erasure of N qubits in Sect. 4.1.

Appendix C: Explicit Construction of a Measurement Dilation for a Maxwell Instrument

Let a Maxwell instrument of the form (11) be given, i. e.,

$$\mathfrak{F}(n)(\rho) = U_n P_n \rho P_n U_n^*, \quad n \in \mathcal{N}. \tag{C1}$$

Following [15] we want to explicitly construct a measurement dilation of $\mathfrak{F}$ of the form (17). To this end we choose $\mathcal{K} = \mathbb{C}^{\mathcal{N}}$ and an orthonormal basis $|n\rangle_{n \in \mathcal{N}}$ in $\mathcal{K}$. Let $\phi \in \mathcal{K}$ be one of these basis vectors, say, $\phi = |1\rangle$.

Further, let $\check{P}_n$ denote the eigenspace of the projector P_n corresponding to the eigenvalues 1 and $|ni\rangle_{i=1,\ldots,\dim \check{P}_n}$ some orthonormal bases in $\check{P}_n$ such that

$$P_n = \sum_i |ni\rangle\langle ni|, \quad \text{for all } n \in \mathcal{N}. \tag{C2}$$

Moreover, let $\check{Q}_n \equiv \mathcal{H} \otimes |n\rangle$ and $Q_n = |n\rangle\langle n|$ denote the projector onto the one-dimensional subspace spanned by $|n\rangle$ for all $n \in \mathcal{N}$. We define a linear map $V_1 : \check{Q}_1 \to \mathcal{H} \otimes \mathcal{K}$ by

$$V_1|ni1\rangle \equiv V_1(|ni\rangle \otimes |1\rangle) \equiv U_n|ni\rangle \otimes |n\rangle = (U_n \otimes \mathbb{1})|nin\rangle \tag{C3}$$

where $i = 1, \ldots, \dim \check{P}_n$ and $n \in \mathcal{N}$.

Lemma 1 *The map $V_1 : \check{Q}_1 \to \mathcal{H} \otimes \mathcal{K}$ is a partial isometry, i. e., satisfies $V_1^* V_1 = \mathbb{1}_{\check{Q}_1}$.*

Proof Let $|mj1\rangle$ and $|ni1\rangle$ be two arbitrary vectors of the orthonormal basis of $\check{Q}_1$ obtained from the orthonormal basis of $\mathcal{H}$ considered above. Then we conclude

$$\langle mj1|V_1^* V_1|ni1\rangle \overset{(C3)}{=} \langle mjm|(U_m^* \otimes \mathbb{1})(U_n \otimes \mathbb{1})|nin\rangle \tag{C4}$$

$$= \langle mj|U_m^* U_n|ni\rangle \underbrace{\langle m|n\rangle}_{\delta_{mn}} \tag{C5}$$

$$= \langle nj|U_n^* U_n|ni\rangle \delta_{mn} \tag{C6}$$

$$= \langle nj|\mathbb{1}|ni\rangle \delta_{mn} \tag{C7}$$

$$= \delta_{ij} \delta_{mn} \tag{C8}$$

$$= \langle mj1|\mathbb{1}_{\check{Q}_1}|ni1\rangle, \tag{C9}$$

which completes the proof of Lemma 1. $\qquad\square$

Next we extend the partial isometry V_1 to a unitary operator $V : \mathcal{H} \otimes \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$. This completes the definition of the quantities $\mathcal{K}, \phi, V, Q$ required for the measurement dilation. It remains to show that $\mathfrak{F} = \mathfrak{D}_{\mathcal{K},\phi,V,Q}$. To this end we write

$$\rho = \sum_{\ell,i,m,j} |\ell i\rangle\langle \ell i|\rho|mj\rangle\langle mj| \tag{C10}$$

and

$$\rho \otimes P_\phi = \sum_{\ell,i,m,j} |\ell i1\rangle\langle \ell i|\rho|mj\rangle\langle mj1| \,. \tag{C11}$$

Further,

$$
\begin{aligned}
&V\big(\rho \otimes P_\phi\big)V^* \\
&\overset{(C11)}{=} \sum_{\ell,i,m,j} V|\ell i1\rangle\langle \ell i|\rho|mj\rangle\langle mj1|\,V^*
\end{aligned}
\tag{C12}
$$

$$
\overset{(C3)}{=} \sum_{\ell,i,m,j} \big(U_\ell \otimes \mathbb{1}\big)|\ell i\ell\rangle\langle \ell i|\rho|mj\rangle\langle mjm|\big(U_m^* \otimes \mathbb{1}\big)
\tag{C13}
$$

$$
= \sum_{\ell,i,m,j} \big(U_\ell|\ell i\rangle\langle \ell i|\rho|mj\rangle\langle mj|U_m^*\big) \otimes |\ell\rangle\langle m| \,.
\tag{C14}
$$

Using

$$Q_n|\ell\rangle\langle m|Q_n = \delta_{\ell n}\,\delta_{mn}\,Q_n \,, \tag{C15}$$

([C14](#)) implies

$$
\begin{aligned}
&\big(\mathbb{1} \otimes Q_n\big)V\big(\rho \otimes P_\phi\big)V^*\big(\mathbb{1} \otimes Q_n\big) \\
&= \sum_{i,j} \big(U_n|ni\rangle\langle ni|\rho|nj\rangle\langle nj|U_n^*\big) \otimes Q_n \,,
\end{aligned}
\tag{C16}
$$

and

$$
\begin{aligned}
&\mathcal{D}_{\mathcal{K},\phi,V,Q}(\rho) \\
&= \mathrm{Tr}_{\mathcal{K}}\big(\big(\mathbb{1} \otimes Q_n\big)V\big(\rho \otimes P_\phi\big)V^*\big(\mathbb{1} \otimes Q_n\big)\big) \\
&\overset{(C16)}{=} \sum_{i,j} U_n|ni\rangle\langle ni|\rho|nj\rangle\langle nj|U_n^* \,,
\end{aligned}
\tag{C17}
$$

since $\mathrm{Tr}\,Q_n = 1$ for all $n \in \mathcal{N}$. The latter expression equals

$$
\begin{aligned}
\mathfrak{F}(n)(\rho) &= U_n P_n \,\rho\, P_n U_n^* \\
&\overset{(C2)}{=} \sum_{i,j} U_n|ni\rangle\langle ni|\rho|nj\rangle\langle nj|U_n^* \,,
\end{aligned}
\tag{C18}
$$

thereby proving that the above construction is a correct measurement dilation of $\mathfrak{F}$.

Next we calculate the reduction of the final state to the demon subsystem and obtain

$$\rho_2 \equiv \text{Tr}_{\mathcal{H}} \sum_{n \in \mathcal{N}} \left(\mathbb{1} \otimes Q_n\right) V\left(\rho \otimes P_\phi\right) V^*\left(\mathbb{1} \otimes Q_n\right)$$
$$\overset{(C16)}{=} \sum_{n \in \mathcal{N}} \text{Tr}\left(P_n \rho P_n\right) Q_n \overset{(B1)}{=} \sum_{n \in \mathcal{N}} p_n Q_n \,. \tag{C19}$$

The corresponding entropy amounts to the Shannon entropy

$$S_2 = S(\rho_2) = -\sum_{n \in \mathcal{N}} p_n \log p_n \overset{(59)}{=} H(p) \geq 0 \,. \tag{C20}$$

In connection with the Szilard principle the following result is interesting:

Proposition 3 *The total entropy of the composed state after the measurement dilation $\mathfrak{D}_{\mathcal{K},\phi,V,Q}$ constructed above exceeds (or equals) the entropy of the state after the corresponding Lüders operation,*

$$\widetilde{S}_1 \leq S_1 + S_2 \,. \tag{C21}$$

Proof of Proposition 3 With the definitions (B1)–(B3) we conclude from the concavity of the von Neumann entropy, see (11.86) in [15],

$$\sum_{n \in \mathcal{N}} p_n S(\rho_n) \overset{(B4)}{=} \sum_{n \in \mathcal{N}} p_n S(\widetilde{\rho}_n) \leq S\left(\sum_{n \in \mathcal{N}} p_n \widetilde{\rho}_n\right) = S_1 \,. \tag{C22}$$

This further implies

$$\widetilde{S}_1 - S_2 \overset{(C20)}{=} \widetilde{S}_1 - H(p) \overset{(B5)}{=} \sum_{n \in \mathcal{N}} p_n S(\rho_n) \overset{(C22)}{\leq} S_1 \,, \tag{C23}$$

and (C21) immediately follows. $\qquad\qquad\square$

Appendix D: The Szilard Engine Revisited

We will reconsider the Szilard engine, but in contrast to the simplified model in Sect. 4.2, adopt a more realistic description of the one-molecule gas and the isothermal expansion after position measurement. In doing so, we will stick to [9] as far as possible, but emphasize the differences to the present approach.

In classical thermodynamics there are various equivalent formulations of the 2nd law including the impossibility of a perpetuum mobile of the second kind. This is a cyclic process transforming heat completely into work without further changes of the environment. The Szilard engine is designed as a possible realization of such a perpetuum mobile but in the present paper we will concentrate on the entropy balance, against the grain, so to speak.

The one-molecule gas is initially confined to a cylindrical box $\mathcal{V}$ with volume V that will be separated into two chambers $\mathcal{V}_R$ and $\mathcal{V}_L$ with equal volumes V/2 by the adiabatic insertion of a piston. Contrary to [9] we will neglect the preparatory process of insertion of the piston since it is only needed for a cyclic process but would be irrelevant for the entropy balance. The Hilbert space of the gas will be chosen as

$$\mathcal{H}_g = \mathcal{H}_R \oplus \mathcal{H}_L \equiv L^2(\mathcal{V}_R) \oplus L^2(\mathcal{V}_L) .\tag{D1}$$

The isothermal expansion cannot be described by a unitary operator acting only on $\mathcal{H}_g$. Thus we have to extend the object system by a heat bath with Hilbert space $\mathcal{H}_b$ and take the Hilbert space of the object system as

$$\mathcal{H} = \mathcal{H}_g \otimes \mathcal{H}_b .\tag{D2}$$

We note that these Hilbert spaces are infinite-dimensional. Strictly speaking, we are restricted to the finite-dimensional case according to the general assumption in Sect. 2 but we do not expect that this will cause any problem.

Initially the state of the object system is assumed to be given by the product state

$$\rho_0 = \rho_g \otimes \rho_b \equiv \frac{1}{2}(\rho_R + \rho_L) \otimes \rho_b \tag{D3}$$

where ρ_R, ρ_L and ρ_b are Gibbs states with the same temperature T corresponding to suitable Hamiltonians. The Hamiltonian for the gas is the one-particle kinetic energy with the boundary condition of vanishing wave functions at the boundaries of $\mathcal{V}_R$ and $\mathcal{V}_L$. Due to symmetry considerations we will assume

$$S(\rho_R) = S(\rho_L) .\tag{D4}$$

The projectors of the first Lüders measurement will be P_R and P_L corresponding to projections onto the subspaces $\mathcal{H}_L$ and $\mathcal{H}_R$, resp., see (D1). These projectors commute with ρ_0 and thus the corresponding total Lüders operation (1) alone would not change the state ρ_0. But we have to perform a conditional action: Depending on the result of this measurement one of two possible isothermal expansions will be performed that are described by unitary operators U_R, $U_L : \mathcal{H} \to \mathcal{H}$. Hence the state of the object system after the conditional action will be

$$\rho_1 = \frac{1}{2}\left(U_R(\rho_R \otimes \rho_b)U_R^* + U_L(\rho_L \otimes \rho_b)U_L^*\right) .\tag{D5}$$

One expects from physical reasons that after the isothermal expansion one would obtain a one-dimensional gas filling the box $\mathcal{V}$ in thermal equilibrium with the heat bath. Hence both density operators in (D5) will be equal to a Gibbs state of the form $\rho_g \otimes \rho_b$, but with a slightly lower temperature than T. However, we will not need this strong thermalization assumption but only the weaker one that can be justified by symmetry considerations:

$$U_R(\rho_R \otimes \rho_b)U_R^* \approx U_L(\rho_L \otimes \rho_b)U_L^* \approx \rho_1 ,\tag{D6}$$

where the second approximation follows from (D5). Eq. (D6) implies

$$S(\rho_1) \approx S\big(U_R\big(\rho_R \otimes \rho_b\big)U_R^*\big) \tag{D7}$$

$$= S\big(\rho_R \otimes \rho_b\big) \tag{D8}$$

$$= S(\rho_R) + S(\rho_b) . \tag{D9}$$

This further gives the following result for the entropy decrease due to the conditional action:

$$S(\rho_1) - S(\rho_0) \overset{(D9,D3)}{\approx} \big(S(\rho_R) + S(\rho_b)\big) - \big(S(\rho_g) + S(\rho_b)\big) \tag{D10}$$

$$= S(\rho_R) - S(\rho_g) \tag{D11}$$

$$\approx -\log 2 . \tag{D12}$$

The last approximation (D12) follows from

$$S(\rho_g) \overset{(D3)}{=} S\Big(\frac{1}{2}\rho_R + \frac{1}{2}\rho_L\Big)$$
$$\overset{(B5)}{=} \frac{1}{2} S(\rho_R) + \frac{1}{2} S(\rho_L) \tag{D13}$$

$$-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} \tag{D14}$$

$$\overset{(D4)}{=} S(\rho_R) + \log 2 . \tag{D15}$$

This entropy decrease has not been calculated directly by Zurek in [9] but follows from his result

$$\Delta A = k_B T \log 2 \tag{D16}$$

for the increase of free energy A of the gas due to the position measurement, see [9] Eq. (20), if we take into account the thermodynamical identity

$$ST = E - A , \tag{D17}$$

and that the intrinsic energy E of the gas does not change due to the measurement. (Note that we have used dimensionless entropy units in this paper and hence set Boltzmann's constant k_B to 1.)

Zurek also considers in [9], section "Measurement by 'quantum Maxwell's demon' ", a measurement dilation similar to that considered in this paper, but only for the pure Lüders measurement, not for the conditional action. Nevertheless, he obtained an entropy increase of $\Delta S = k_B \log 2$ of the demon's state that exactly

compensates the entropy decrease of the gas calculated above, and related this entropy increase to the loss of "mutual information", see [9] Eq. (36). It will be instructive to compare these considerations with the measurement dilation scheme considered in Sect. 3 applied to the Szilard engine model.

We choose the demon's Hilbert space as $\mathcal{K} = \mathbb{C}^2$ with orthonormal basis (r, ℓ) and projectors $Q_R = |r\rangle\langle r|$, $Q_L = |\ell\rangle\langle \ell|$. The initial state of the demon will be chosen as $\phi = r$. Further we choose a unitary operator $V : \mathcal{H}_g \otimes \mathcal{H}_b \otimes \mathcal{K} \to \mathcal{H}_g \otimes \mathcal{H}_b \otimes \mathcal{K}$ satisfying

$$V\big(\psi_R \otimes \psi_b \otimes r\big) = U_R\big(\psi_R \otimes \psi_b\big) \otimes r, \tag{D18}$$

$$V\big(\psi_L \otimes \psi_b \otimes r\big) = U_L\big(\psi_L \otimes \psi_b\big) \otimes \ell, \tag{D19}$$

for all $\psi_R \in \mathcal{H}_R$, $\psi_L \in \mathcal{H}_L$, and $\psi \in \mathcal{H}_b$.

The factors of the initial state $\rho_0 = \rho_g \otimes \rho_b$ will have spectral decompositions of the following form

$$\rho_g = \frac{1}{2}\big(\rho_R + \rho_L\big) = \frac{1}{2}\sum_i p_i\Big(|\psi_R^{(i)}\rangle\langle\psi_R^{(i)}| + |\psi_L^{(i)}\rangle\langle\psi_L^{(i)}|\Big) \tag{D20}$$

$$\rho_b = \sum_j b_j |\psi_b^{(j)}\rangle\langle\psi_b^{(j)}|, \tag{D21}$$

where we have used that the eigenvalues p_i of ρ_R are the same as those for ρ_L due to symmetry.

After a straight forward calculation using (D20) and (D21) we obtain for the total state ρ_{12} after the interaction the expected result

$$\rho_{12} = V\big(\rho_g \otimes \rho_b \otimes |r\rangle\langle r|\big)V^* = \rho_1 \otimes \frac{1}{2}(|r\rangle\langle r| + |\ell\rangle\langle\ell|), \tag{D22}$$

with ρ_1 according to (D6). Since ρ_{12} commutes with $\mathbb{1} \otimes Q_R$ and $\mathbb{1} \otimes Q_L$ the final Lüders measurement does not change this state:

$$\rho_{12} = \big(\mathbb{1} \otimes Q_R\big)\rho_{12}\big(\mathbb{1} \otimes Q_R\big) + \big(\mathbb{1} \otimes Q_L\big)\rho_{12}\big(\mathbb{1} \otimes Q_L\big), \tag{D23}$$

analogously to the measurement dilation considered in [9]. The difference to our calculation is that we have no correlation between object system and demon in the final state ρ_{12} and the separation into partial traces considered in Sect. 2 is superfluous.

Consequently, the total entropy during the conditional action will be constant since the entropy of the demon increases by $\Delta S_d = \log 2$ as can be directly read off the final demon's state in (D22), and the entropy of the object system decreases according to $S(\rho_0) - S(\rho_1) = -\log 2$, see (D11) and (D15).

We should add a remark on the role of approximations in the present problem of Szilard's engine. These approximations simplify the presentation but are not crucial for the total entropy balance that is guaranteed by the measurement dilation as explained in Sect. 3. For example, if we cancel the approximation, that the

isothermal expansion reaches the same state in both cases, see (D6), then in the final state ρ_{12} after the interaction a small correlation would remain. The following measurement and the separation of the states of subsystems would lead to a small further increase of entropy without changing the final result substantially. A variant of the Szilard engine without any need for approximations would be obtained by replacing the final isothermal expansion by an adiabatic expansion without any heat bath. Of course this runs counter to the original motive of constructing a cyclic heat engine.

References

1. Leff, H., Rex, A. (eds.): Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing. Institute of Physics, Bristol (2003)
2. Earman, J., Norton, J.D.: Exorcist XIV: The Wrath of Maxwell's Demon. Part I. From Maxwell to Szilard. Stud. Hist. Philos. Mod. Phys. **29**(4), 435–471 (1998)
3. Earman, J., Norton, J.D.: Exorcist XIV: The Wrath of Maxwell's Demon. Part II. From Szilard to Landauer and Beyond. Stud. Hist. Philos. Mod. Phys. **30**(1), 1–40 (1999)
4. Szilard, L.: Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen (On the reduction of entropy in a thermodynamic system by the intervention of intelligent beings). ZS. f. Phys. **53**(11–12), 840–856 (1929)
5. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**, 183–191 (1961)
6. Bennett, C.H.: The thermodynamics of computation-a review. Int. J. Theor. Phys. **21**(12), 905–940 (1982)
7. Leff, H.S., Rex, A.F.: Entropy of measurement and Erasure: Szilard's membrane model revisited. Am. J. Phys. **62**, 994–1000 (1994)
8. Norton, J.D.: A hot mess. Inference: International Review of Science **4** (4), (2019)
9. Zurek, W.H.: Maxwell's demon, Szilard's engine and quantum measurements. In: Moore, G.T., Seully, M.O. (eds.) Front. Nonequilib. Stat. Mech., pp. 151–161. Plenum Press, New York (1984)
10. Lubkin, E.: Keeping the entropy of measurement: Szilard revisited. Int. J. Theor. Phys. **26**, 523–535 (1987)
11. Lloyd, S.: Use of mutual information to decrease entropy: implications for the second law of thermodynamics. Phys. Rev. A **39**, 5378–5386 (1987)
12. Biedenharn, L.C., Solem, J.C.: A quantum-mechanical treatment of Szilard's engine: implications for entropy of information. Found. Phys. **25**, 1221–1229 (1995)
13. Lloyd, S.: Quantum-mechanical Maxwell's demon. Phys. Rev. A **56**, 3374–3382 (1997)

14. Vedral, V.: Landauer's erasure, error correction and entanglement. Proc. R. Soc. Lond. A **456**, 969–984 (2000)
15. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
16. Sagawa, T., Ueda, M.: Generalized Jarzynski equality under nonequilibrium feedback control. Phys. Rev. Lett. **104**, 090602 (2010)
17. Shiraishi, N., Sagawa, T.: Fluctuation theorem for partially masked nonequilibrium dynamics. Phys. Rev. E **91**, 012130 (2015)
18. Busch, P., Lahti, P.J., Pellonpää, J.-P., Ylinen, K.: Quantum Measurement. Springer, Berlin (2016)
19. von Neumann, J.: Mathematische Grundlagen der Quantenmechanik, Springer-Verlag, Berlin, 1932, English translation: Mathematical Foundations of Quantummechanics. Princeton University Press, Princeton (1955)
20. Schmidt, H.-J., Gemmer, J.: Sequential measurements and entropy, Preprint quant-ph/2001.04400
21. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
22. Lindblad, G.: Entropy, information and quantum measurements. Commun. Math. Phys. **33**, 305–322 (1973)
23. Busch, P., Lahti, P.J., Mittelstädt, P.: The Quantum Theory of Measurement, 2nd revised edn. Springer, Berlin (1996)
24. Edwards, C.M.: The theory of pure operations. Commun. Math. Phys. **24**, 260–288 (1972)