Reversible dynamics in strongly non-local Boxworld systems

Sabri W Al-Safi and Anthony J Short

1 DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
2 H H Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK

E-mail: S.W.Al-Safi@damtp.cam.ac.uk and tony.short@bristol.ac.uk

Received 4 February 2014, revised 26 May 2014
Accepted for publication 12 June 2014
Published 29 July 2014

Abstract
In order to better understand the structure of quantum theory, or speculate about theories that may supersede it, it can be helpful to consider alternative physical theories. ‘Boxworld’ describes one such theory, in which all non-signaling correlations are achievable. In a limited class of multipartite Boxworld systems—wherein all subsystems are identical and all measurements have the same number of outcomes—it has been demonstrated that the set of reversible dynamics is ‘trivial’, generated solely by local relabellings of measurement choices and outcomes, and permutations of subsystems. We develop the convex formalism of Boxworld to give an alternative proof of this result, then extend this proof to all multipartite Boxworld systems, and discuss the potential relevance to other theories. These results lend further support to the idea that the rich reversible dynamics in quantum theory may be the key to understanding its structure and its informational capabilities.

Keywords: Boxworld, reversibility, nonlocality

1. Introduction

To gain a better understanding of quantum theory, and to explore possible future modifications, it can be helpful to view quantum theory from the ‘outside’—as one member of a broader class of probabilistic physical theories. One such approach is to consider the class of
general probabilistic theories [1–4], which are based on operational notions that allow many different theories to be represented using the same intuitive mathematical formalism. A natural alternative to quantum theory within this class is Boxworld [5] (originally called generalized non-signalling theory in [1]). Like quantum theory, this theory admits non-local correlations which cannot be explained by any locally realistic model [6]. In fact it admits all non-signaling correlations, including those which maximally violate Bell inequalities, such as the well-known PR-Box [7]. These super-strong correlations cause Boxworld to differ markedly from the world we observe: for example, any distributed computation could be performed with the transmission of a single bit [8], and bit-commitment would be possible without using relativistic effects [9].

In attempting to better understand what makes quantum theory uniquely successful at describing the world, attention has been given to various principles which we expect nature to obey, but which are often violated by other probabilistic theories. A common example of this is reversibility (or transitivity) [2, 10–13], which demands that any two pure states are linked by a reversible transformation.

In [14], Colbeck et al proved that, if one assumes that all systems are identical and all measurements have the same number of outcomes, the set of reversible Boxworld multipartite dynamics is generated by local operations and permutations of systems. Such Boxworld systems cannot experience reversible interactions and thus violate reversibility. In particular, a PR-box (or any other entangled state) could not generated reversibly from an initial product state.

Other recent results have also highlighted the importance of reversibility. It has been shown that two systems whose state-spaces are $d$-dimensional balls can only interact in a continuous and reversible way if $d = 3$ (in which case the systems correspond to qubits in the Bloch-sphere representation) [15]. Furthermore, any theory in which local systems are identical to qubits and in which there exists at least one continuous reversible interaction must globally be identical to quantum theory [10]. Generalizations of such results may have great significance in explaining why our world looks quantum, or in finding theories which may potentially supersede quantum theory.

A common feature of the reversibility results cited above [10, 14, 15] is that they consider interactions between systems which are locally identical. However, the reversibility of quantum theory carries over to the case where, for example, the systems have differing Hilbert space dimension. Could there exist non-trivial reversible interactions between different types of system in these other general probabilistic theories? In this paper we provide an alternative proof of Colbeck et al’s result and then extend it to apply to any combination of Boxworld systems, including different types (so long as none of the systems are classical).

The structure of the paper is as follows: in section 2 we outline the convex formalism of Boxworld and introduce some extra terminology useful for our exposition. In section 3 we prove some results about the structure of this convex set, which we show in section 4 are sufficient to recover the result of [14]. We then extend this proof to the general case of non-identical systems.

2. Set-up and notation

In order to compare quantum theory to alternative theories, such as Boxworld, it is helpful to define a mathematical framework which is broad enough to describe any such theory. Here we consider an operational framework for defining general probabilistic theories [1, 2] in which the state of a system is specified by the probabilities it assigns to effects, or
measurement outcomes. Moreover, for each system we will assume there exists a finite set of fiducial measurements, which are sufficient to deduce the outcome probabilities of any other measurement. For example, in quantum theory the Pauli measurements $\sigma_x$, $\sigma_y$ and $\sigma_z$ form a set of fiducial measurements for a qubit system, although there are infinitely many such possibilities. Boxworld systems differ in that one begins with a fixed, finite set of fiducial measurements, and then imposes restrictions on the achievable outcome statistics for those measurements.

Consider a composite system composed of $N$ individual subsystems. One possible global measurement on the joint system involves a local fiducial measurement being performed on each subsystem. The value $p(a_1, \ldots, a_N|x_1, \ldots, x_N)$ then denotes the probability of outcomes $a_1, \ldots, a_N$ occurring at systems $1, \ldots, N$ respectively, given that the measurement choices $x_1, \ldots, x_N$ were made on those systems. We assume that a complete specification of the values $p(a_1, \ldots, a_N|x_1, \ldots, x_N)$ is sufficient to determine the state of the joint system (an assumption commonly known as local tomography [1, 2]). These values obey the normal laws of probability:

$$0 \leq p(a_1, \ldots, a_N|x_1, \ldots, x_N) \leq 1,$$

and for any fixed choice of the $x_i$:

$$\sum_{a_1,\ldots,a_N} p(a_1, \ldots, a_N|x_1, \ldots, x_N) = 1.$$

To ensure that information cannot be sent between the systems (since, for example, they may be spatially separated), we demand also that the no-signaling condition is satisfied by the joint distribution $p$. This condition says that the outcome statistics for any subset $\Omega$ of the $N$ systems must not be affected by measurement choices made on systems not in $\Omega$, i.e. there is no way for one system to signal information to any other systems. In mathematical terms, we demand that for all $\Omega \subset [N]$, the marginal distribution

$$\sum_{a_1,\ldots,a_N} p(a_1, \ldots, a_N|x_1, \ldots, x_N)$$

is well-defined, independent of the value of $x_i$ for systems $i \notin \Omega$.

**Boxworld** consists of all multipartite states whose joint outcome distributions for the fiducial measurements obey (1, 2, 3). Even for single systems this allows for many more states than are possible in quantum theory. For example, it is impossible for a qubit to be in a state which gives a deterministic outcome for every Pauli measurement, but there exist Boxworld states which give deterministic outcomes for every fiducial measurement. This leniency regarding states results in strong restrictions on the possible measurements: single Boxworld systems have only a single set of fiducial measurements, and multipartite Boxworld systems do not allow for entangled measurements [5].

Instead of working directly with conditional probability distributions $p(a_1, \ldots, a_N|x_1, \ldots, x_N)$, it is often convenient to represent general probabilistic theories using real vector spaces, so that states $s$ and effects $e$ are specified by vectors such that $\langle e, s \rangle$ equals the probability of effect $e$ occurring for a system in state $s$. For a multipartite Boxworld system comprising $N$ subsystems, suppose there are $M^{(i)}$ fiducial measurement choices on system $i$, and $K^{(j)}$ outcomes for the $j$th fiducial measurement on system $i$. When $M^{(i)} = 1$, a single probability distribution is sufficient to describe the state of the system, and we say the system is classical (the results in this paper apply exclusively to the case where all systems are non-classical).
We now give a procedure for constructing the local fiducial effect vectors for system $i$ of our multipartite Boxworld system. First, pick a linearly independent set of vectors $\{\mathcal{U}^{(i)}, X^{(i)}_{a_1|x_1}, \ldots, X^{(i)}_{a_d|x_d}\} \subset \mathbb{R}^d$ for $1 \leq x_i \leq M^{(i)}$ and $1 \leq a_i \leq K^{(i)}_i - 1$, where $d = 1 + \sum_{x_i = 1}^{M^{(i)}} (K^{(i)}_i - 1)$. $\mathcal{U}^{(i)}$ represents the unit effect: the unique effect for which any allowed (normalized) state gives probability 1. $X^{(i)}_{a_i|x_i}$ is the fiducial effect corresponding to measuring $x_i$ and obtaining outcome $a_i$. The remaining fiducial effect vectors are defined $X^{(i)}_{iK_{i+1}|x_i} = \mathcal{U}^{(i)} - \sum_{a_i|x_i} P^{(i)}_{a_i|x_i} X^{(i)}_{a_i|x_i}$.

Note that the linear independence of the set $\{\mathcal{U}^{(i)}, X^{(i)}_{a_i|x_i}\}$ implies that any conditional probability distribution $p(a_i|x_i)$ over the outcomes of the fiducial measurements corresponds to a unique vector $s$ defined by $\langle X^{(i)}_{a_i|x_i}, s \rangle = p(a_i|x_i)$. Thus the set of vectors which has positive inner product with all fiducial effect vectors (and inner product 1 with $\mathcal{U}^{(i)}$) forms the Boxworld state space for system $i$. Conversely, any vector which has a positive inner product less than or equal to 1 with every state $s$ corresponds to an effect that can in principle be measured. For example, for a fixed fiducial measurement $x_i$ with distinct outcomes $a_i$ and $a_j$, the effect $X^{(i)}_{a_i|x_i} + X^{(i)}_{a_j|x_i}$ corresponds to the process of measuring $x_i$, and either outcome $a_i$ or $a_j$ occurring (this may be physically realized by ‘forgetting’ which of the two outcomes occurred).

It turns out that the tensor product of the vector spaces characterizing each individual system provides a neat representation of states and effects in joint Boxworld systems [1]. Suppose system $i$ is represented by a real vector space $\mathbb{R}_i$ as above, and let $\mathbb{R}_N = \mathbb{R}_1 \otimes \ldots \otimes \mathbb{R}_N$. The $N$-partite fiducial effects are defined to be the vectors of the form $X^{(1)}_{a_1|x_1} \otimes \ldots \otimes X^{(N)}_{a_N|x_N}$, where $X^{(i)}_{a_i|x_i}$ is a fiducial effect on system $i$. The $N$-partite unit effect is defined by $\mathcal{U} = \mathcal{U}^{(1)} \otimes \ldots \otimes \mathcal{U}^{(N)}$.

Any Boxworld state whose measurement statistics obey (1, 2, 3) corresponds to a unique vector $s \in \mathbb{R}_N$, such that $\langle \mathcal{U}, s \rangle = 1$ (i.e. the state is normalized) and $p(a_1, \ldots, a_N|x_1, \ldots, x_N) = \langle X^{(1)}_{a_1|x_1} \otimes \ldots \otimes X^{(N)}_{a_N|x_N}, s \rangle$ [14]. Let the set of allowed state vectors be denoted by $\mathcal{S} \subset \mathbb{R}_N$. Pure product states are those of the form $s^{(1)} \otimes \ldots \otimes s^{(N)}$ where $s^{(i)}$ deterministically assigns to each fiducial measurement $x_i$ on system $i$, a definite outcome $1 \leq s^{(i)}_a \leq K^{(i)}_i$, i.e. $X^{(i)}_{a_i|x_i} s^{(i)} = 1$ iff $s^{(i)}_a = a_i$.

Note that in quantum theory, the Pauli measurements are not the only measurements that can be made on a qubit system, despite constituting a fiducial set of measurements. Indeed, any positive semi-definite operator $P \leq I$ corresponds to an outcome of some carefully constructed measurement. Similarly, in Boxworld the fiducial measurements are not the only measurements that can be performed, and the fiducial effects are not the only vectors which correspond to measurement outcomes. However, since the fiducial measurements fully characterize a state $s$, the probability of obtaining any non-fiducial effect must be obtained from the values $p(a_1, \ldots, a_N|x_1, \ldots, x_N)$ associated with $s$. In fact, it can be shown that a function which maps every state to a probability in the interval $[0, 1]$, and behaves sensibly with respect to probabilistic mixtures of states, can be realized by a vector $e$ such that $\langle e, s \rangle$ is the outcome of that function [1, 2].

Consequently, the set of allowed $N$-partite effects in Boxworld is the set of vectors $e \in \mathbb{R}_N$ such that $\langle e, s \rangle \in [0, 1]$ for all $s \in \mathcal{S}$. This includes the $N$-partite fiducial effects, as well as linear combinations of the fiducial effects that are positive and sub-normalized for all states. As for single systems, these non-fiducial effects can in principle be measured by ignoring particular measurement outcomes, forgetting which of several outcomes occurred, or performing measurements probabilistically. If $\langle e, s \rangle = 1$, we will say that the state $s$ hits the effect $e$. We will be particularly interested in effects of the form $E = \sum_{e} e_{a_i}$, where each $e_i$ is a fiducial effect. We say in this case that $\{e_{a_i}\}$ forms a decomposition of $E$, or $E$ admits the
decomposition \( \{ e_i \} \). We will tend to use lowercase letters for fiducial effects, and uppercase for sums of extreme ray effects.

An effect \( E \) is **multiform** if it can be written \( E = \sum e_i = \sum f_\beta \) where \( \{ e_i \} \) and \( \{ f_\beta \} \) are distinct sets of fiducial effects. Effects of the form \( X^{(1)}_{a|x_1} \otimes \cdots \otimes \mathcal{U}^{(i)} \otimes \cdots \otimes X^{(N)}_{a|x_N} \), where exactly one component of the tensor product is the unit effect, and the remainder are fiducial effects, are said to be **sub-unit** effects, or an **i-sub-unit** effect if the \( i \)th component is the unit effect. For each \( x_i \), \( \mathcal{U}^{(i)} \) has a distinct decomposition \( \sum_{a=1}^{K} X^{(i)}_{a|x_i} \), hence sub-unit effects are trivially multiform (as we assume all systems are non-classical and hence have at least two fiducial measurements).

Fiducial and sub-unit effects are tensor products of vectors, so it makes sense to refer to their \( i \)th component, e.g. \( E^{(i)} = X^{(i)}_{a|x_i} \). For a subset \( \Omega \subseteq [N] \) we will write \( E^{\Omega} = \bigotimes_{\alpha \in \Omega} E^{(\alpha)} \), e.g. \( E^{\{1,3\}} = X^{(1)}_{a|x_1} \otimes X^{(3)}_{a|x_3} \).

Finally, we say that a set of fiducial effects \( \{ e_i \} \) **strictly covers** the effect \( E \) if there is some (strict) subset \( B \subset A \) such that \( \sum_{a \in B} e_a = E \).

### 3. Decompositions

We now prove some results concerning multiform effects. Given that none of the systems are classical, the simplest multiform effects are the sub-unit effects, which have various decompositions according to the different measurement choices on the system whose component is the unit effect. The following Lemma shows that these are the only possible decompositions of a sub-unit effect.

**Lemma 1.** Let \( E = \sum e_i \) be an i-sub-unit effect. Then each fiducial effect \( e_i \) satisfies \( e^{(j)}_i = E^{(j)} \) for all components \( j \neq i \). Moreover, the set of \( i \)th components \( \{ e^{(i)} \} \) forms a fiducial measurement on system \( i \).

**Proof.** See appendix A

**Corollary 1.** Let \( E = \sum_{a=1}^{m} e_a \) be an i-sub-unit effect. Then \( \{ e_a \} \) does not strictly cover a multiform effect.

**Proof.** From Lemma 1 it follows that \( e^{(j)}_i = E^{(j)} \) for all components \( j \neq i \), and that \( \sum e^{(i)}_a = \mathcal{U}^{(i)} \); in other words, the decompositions of \( E \) are in a one-to-one correspondence with the local fiducial effect decompositions of \( \mathcal{U}^{(i)} \). Due to the linear relations between the local fiducial effect vectors defined in section 2, it can be verified that the only local decompositions of \( \mathcal{U}^{(i)} \) are those obtained by fixing a measurement choice \( x_i \) and summing all the fiducial effect vectors which correspond to an outcome for that measurement:

\[
\mathcal{U}^{(i)} = \sum_{a} X^{(i)}_{a|x_i}.
\] (4)

In particular, there are only a finite number of such decompositions, and the sets of local fiducial effects making up these decompositions are pairwise disjoint. Likewise, there are only a finite number of decompositions of \( E \), and the sets of composite fiducial effects making up these decompositions are pairwise disjoint.

Suppose that \( \{ e_i \} \) covers a multiform effect, i.e. (after relabelling the \( e_i \) if necessary) for some integer \( s < r \) and some integer \( t \), there exist distinct sets \( \{ e_a \}^t_{a=1} \), \( \{ f_\beta \}^s_{\beta=1} \) of fiducial effects such that \( \sum_{a=1}^{t} e_a = \sum_{\beta=1}^{s} f_\beta \). Then \( \{ f_1, \ldots, f_s, e_{s+1}, \ldots e_r \} \) is a decomposition of \( E \).
distinct from \{e_1, \ldots, e_r\}. However, both these sets contain \(e_n\), contradicting the fact that the decompositions of \(E\) are pairwise disjoint. Therefore \(\{e_n\}\) does not cover a multiform effect.

For convenience we will assume from here on that the systems are arranged in order of increasing numbers of measurement outcomes, i.e. \(K_1^{(i)} \leq K_i^{(j)}\) and \(K_1^{(j)} \leq K_i^{(j+1)}\); this amounts to no more than a relabelling of systems and measurement choices. \(K_1^{(1)}\) is therefore the smallest number of outcomes possible for any fiducial measurement. The following lemma restricts the type of effects which can have small decompositions.

**Lemma 2.** For \(r \leq K_1^{(1)}\) suppose that \(\{e_1\}_{\alpha=1}^r\) does not cover any sub-unit effects. Then for any fiducial effect \(f \notin \{e_1\}\), there is a pure product state which hits \(f\) but none of the \(e_\alpha\).

**Proof.** See appendix B

**Corollary 2.** The only multiform effects which have a decomposition with exactly \(K_1^{(1)}\) elements are sub-unit effects.

**Proof.** Suppose \(E = \sum_{\alpha=1}^r e_\alpha = \sum_{\beta=1}^s f_\beta\) are distinct decompositions, with \(r = K_1^{(1)}\), and suppose without loss of generality that \(f_1 \notin \{e_\alpha\}_{\alpha=1}^r\). Every pure product state which hits \(f_1\) must also hit one of the \(e_\alpha\), so it follows from lemma 2 that \(\{e_\alpha\}\) covers a sub-unit effect. By lemma 1, every decomposition of a sub-unit effect has at least \(K_1^{(1)}\) elements, hence \(E\) is itself a sub-unit effect.

4. Dynamics

As well as states and measurements, a physical theory must have some notion of dynamics which transform the state of a system. We call \(T\) an allowed transformation if it is a convex-linear mapping of \(\Sigma\) to itself. i.e. for \(s_1, s_2 \in \Sigma\) and \(0 \leq p \leq 1\), \(T(ps_1 + (1-p)s_2) = pT(s_1) + (1-p)T(s_2)\). Since \(\Sigma\) lies in the non-trivial affine hyperplane \(\{v \in \mathbb{R}_N : \langle U, v \rangle = 1\}\), it can be shown that any map \(T\) which is convex-linear on \(\Sigma\) can be extended to a linear map on \(\mathbb{R}_N\). From here on we will therefore assume that \(T\) is a linear map on \(\mathbb{R}_N\).

Operationally, a transformation is determined by how it affects outcome probabilities. Since \(\langle e, T(s) \rangle = \langle T^\dagger(e), s \rangle\), any transformation \(T\) may equivalently be defined by how its adjoint linear map \(T^\dagger\) acts on effects. From here on we will use the latter perspective (this is similar to working in the Heisenberg picture in quantum theory).

\(T^\dagger\) is reversible if there exists an allowed transformation \(S\) such that \(S T^\dagger(e) = e\) for all effects \(e\). We will assume this is the case from here on and write \(S = T^{-1}\). An interesting property of box-world is that all effects lie in the convex cone whose extreme rays are the fiducial effects. It can easily be checked that if \(T^\dagger\) is reversible, it must map this cone onto itself and hence map fiducial effects to fiducial effects. Moreover if \(E = \sum_{\alpha} e_\alpha\) then \(T^\dagger(E) = \sum_{\alpha} T^\dagger(e_\alpha)\). Clearly, \(T^\dagger\) also maps multiform effects to multiform effects.

4.1. Recovering the identical systems case

Using the results of section 3, it is relatively easy to recover the main result of [14]: that all reversible transformations on Boxworld are trivial as long as all the systems are identical. In fact, we only require that \(K_1^{(i)} = K_1^{(1)}\) for all systems \(i\).
The action of reversible transformations on sub-unit effects forms a crucial step in the proof. Let $E$ be a $i$-sub-unit effect for some $i$, and note that $E$ has a decomposition with exactly $r = K^{(1)}_i$ elements (according to the measurement on system $i$ which has $r$ outcomes). Because system $i$ is non-classical, the local unit effect $U^{(i)}$ on system $i$ can be multiply decomposed into local fiducial effects, with at least one decomposition consisting of $K^{(1)}_i$ elements. Therefore $T^i(E)$ is a multiform composite fiducial effect with at least one decomposition with exactly $K^{(1)}_i$ elements, hence by corollary 2, $T^i(E)$ is a sub-unit effect.

Since reversible transformations permute the set of sub-unit effects, they preserve the property that a pair of fiducial effects differ in only one component. To see this, consider a pair of fiducial effects $e_1$ and $e_2$ which differ only on system $i$: note they each belong to some decomposition of an $i$-sub-unit effect $E$. It follows that $T^i(e_1)$ and $T^i(e_2)$ each belong to some decomposition of the sub-unit effect $T^i(E)$, hence by lemma 1 differ in only one component.

Let $\mathcal{A}_i$ be the set $\{X^{(i)}_{a(1)}\}$ of fiducial effects on system $i$. By viewing the fiducial effect $X^{(i)}_{a(1)} \otimes \cdots \otimes X^{(N)}_{a(N)}$ as a string $X = (X^{(i)}_{a(1)}, \ldots, X^{(N)}_{a(N)})$, $T^i$ can be interpreted as a permutation of the set $\mathcal{A}_i \times \cdots \times \mathcal{A}_n$. The Hamming distance between two strings $X_1, X_2$ is the number of components in which they differ, i.e. $\# \{i: X^{(i)}_1 \neq X^{(i)}_2\}$. By the above comments, $T^i$ preserves a Hamming distance of 1 between pairs of strings. The following lemmas and resultant theorem are proved in [14]. Importantly, the proofs do not assume identical systems or equal numbers of outcomes for local fiducial measurements.

**Lemma 3.** (Proved in [14]). Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be finite alphabets, and $Q$ be a bijective map from $\mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ to itself. If $Q$ preserves a Hamming distance of 1 between pairs of strings, then it is a composition of operations which permute components and local permutations acting on individual components.

**Lemma 4.** (Proved in [14]). The only reversible transformations allowed on single Boxworld systems are relabellings of measurement choices, and relabellings of measurement outcomes.

**Theorem 1.** (Proved in [14]). The only reversible transformations allowed in Boxworld with identical non-classical systems are permutations of systems, and local relabellings of measurement choices and measurement outcomes.

4.2. Extending to the general case

We now relax the condition that $K^{(i)}_i = K^{(1)}_i$ for all systems $i$. This makes the task more complicated, however the method of proof is still to argue that $T^i$ permutes the set of sub-unit effects.

**Lemma 5.** Reversible Boxworld transformations map sub-unit effects to sub-unit effects, as long as none of the systems are classical.

**Proof.** See appendix C for a detailed proof. However, we will sketch the main ideas of the proof here.

If $K^{(i)}_i > K^{(1)}_i$, then sub-unit effects at system $i$ will not have decompositions with $K^{(1)}_i$ elements, so it is not possible to directly apply corollary 2. However, corollary 2 can still be applied for every system $i$ with $K^{(i)}_i = K^{(1)}_i$, such that $T^i$ permutes the set of sub-unit effects within this set of systems.
Now consider a system \( j \) for which \( K_j^{(2)} \) is the next possible greater value than \( K_j^{(1)} \). A \( j \)-sub-unit effect \( E \) must be transformed to something which is multiform with some decomposition \( \{e_i\} \) involving \( K_j^{(1)} \) elements. It turns out that if (a) \( \sum e_i \) is not a sub-unit effect, and (b) \( T^j \) permutes sub-unit effects on systems with \( K_j^{(1)} = K_i^{(1)} \), then for any other decomposition \( \{f_j\} \) of \( T^j(E) \) there exists a pure product state \( s \) which hits \( f_1 \) but none of the effects \( \{e_i\} \).

This sets up an iterative process, at each stage assuming that \( T^j \) permutes the sub-unit effects on systems with smaller numbers of outcomes. The iteration terminates when \( K_j^{(1)} \) takes on its maximal value, and \( T^j \) thus permutes the complete set of sub-unit effects in the multipartite system.

Lemma 5 implies that \( T^j \) preserves a Hamming distance of 1 between effects in the general case, so we can apply lemmas 3 and 4 to deduce that reversible transformations are compositions of permutations of systems and local relabellings of measurement choices and measurement outcomes. However, not every permutation of systems is actually possible. We say that two Boxworld systems are of the same type if they have the same number of fiducial measurement choices, and the fiducial measurements from each system can be matched up so that paired measurements have the same number of outcomes.

**Lemma 6.** Let \( T^j \) be an allowed reversible Boxworld transformation which is a permutation of systems \( P \), followed by a composition of local relabellings \( Q \). Then \( P \) can only permute systems of the same type.

**Proof.** Suppose that \( P \) takes system \( i \) to system \( i' \). Because \( Q \) is constant on any sub-unit effect and \( T^j \) permutes sub-unit effects (by lemma 5), \( P \) also permutes sub-unit effects.

Any two fiducial effects \( e_1, e_2 \) which differ only by measurement outcome on system \( i \), belong to the same decomposition of some \( i \)-sub-unit effect \( E \). \( P(e_1) \) and \( P(e_2) \) therefore belong to the same decomposition of \( P(E) \), which by the above reasoning is an \( i' \)-sub-unit effect for some \( i' \). It follows from lemma 1 that \( e_1 \) and \( e_2 \) differ only by measurement outcome on system \( i' \).

If instead \( e_1, e_2 \) differ by measurement choice on system \( i \), then they belong to distinct decompositions of \( E \) (as above), hence \( P(e_1) \) and \( P(e_2) \) belong to different decompositions of \( P(E) \), therefore differ by measurement choice on system \( i' \).

Hence \( P \) must map the fiducial effects on system \( i \) onto the fiducial effects on system \( i' \) in such a way that effects belong to the same measurement choice after the transformation iff they also did before the transformation. The only way this is possible is if the systems are of the same type.

**Theorem 2.** The only reversible transformations of non-classical systems allowed in Boxworld are permutations of systems of the same type, and local relabellings of measurement choices and measurement outcomes.

**Proof.** Suppose that \( T^j \) is an allowed reversible Boxworld transformation. By lemma 5, \( T^j \) maps sub-unit effects to sub-unit effects, hence \( T^j \) preserves a Hamming distance of 1 when considered as a permutation of strings of local fiducial effects. Therefore by lemmas 3 and 4, \( T^j \) is a composition of permutations of systems and local relabellings of measurement choices and outcomes. These operations commute, so we may assume that the permutation of systems
occurs first, followed by a local relabelling on each system. By lemma 6, \( T' \) may permute only those systems which are of the same type.

To complete the proof, we need only check that all transformations of the above form are allowed. This is obvious from considering the action on distributions
\[
p (a_1, ..., a_N|x_1, ..., x_N).
\]

5. Discussion

We have refined and extended the result of [14], and demonstrated that—as long as no system is classical—reversible dynamics in arbitrary joint Boxworld systems always take the form of permutations of systems of the same type, followed by relabellings of measurement choices and outcomes on individual systems. If any single system is classical, the result immediately fails: any local operation on another system which is conditioned on a particular outcome on the classical system corresponds to an allowed, non-trivial reversible interaction.

This places Boxworld in stark contrast to quantum theory and nature itself, in which systems are clearly able to interact in a reversible way, commonly becoming entangled via continuous and reversible processes. In fact quantum theory obeys a much stronger principle than reversibility: any set of perfectly distinguishable states of a system can be mapped via unitary evolution to any other set of the same size. The importance of reversibility in quantum theory is suggested by the prevalent usage of it (or stronger versions of it) as an axiom in numerous information-theoretic derivations of quantum theory [2, 4, 10, 13, 16, 17].

It is worth comparing our proof with the proof in [14] in a little more detail. In [14] the authors employ a fixed (albeit natural) choice of vectors to represent fiducial effects, for which it turns out that all reversible transformations correspond to orthogonal maps. Since the number of measurement choices \( M \) and outcomes \( K \) are constant across systems in their analysis, the same vector space representation can be chosen for every system, and the fact that \( T' \) preserves inner products between fiducial effect vectors can be exploited. Specifically, the inner products between fiducial effect vectors which differ on only one system take on unique values, so that \( T' \) obeys the conditions of lemma 3 as a permutation on strings of local fiducial effects.

On first impressions, it appears that this method of proof might naturally extend to the case where not all systems are identical. However, it can be shown that the proof breaks down for joint systems with varying values of \( M \) and \( K \) which are carefully chosen to generate redundancies in the values of the inner products between fiducial effect vectors [18]. Indeed, there exist an infinite number of such ‘pathological’ choices of \( M \) and \( K \), and there is no clear way of embedding these pathological cases into larger Boxworld systems in such a way as to inherit their dynamical properties.

Rather than using a fixed representation, our method of proof exploits the linear structure of Boxworld which is inherent in any representation. As a consequence our techniques are more readily applied to the case of non-identical systems, and pave the way to understanding reversible dynamics in other general probabilistic theories. For example, it is interesting to explore whether the proof extends to theories which are ‘maximally non-local’ with respect to their local state spaces, i.e. which admit any composite state consistent with local fiducial measurements [19] (the joint state spaces of such theories are commonly referred to as the ‘maximal tensor product’ of the local state spaces). Such theories share the property of Boxworld that all effects lie in a convex cone whose extreme rays are product effects; consequently, the set of product effects must be permuted by any reversible transformation. However, it turns out that it is possible to construct a pair of local state spaces whose maximal
tensor product does permit non-trivial reversible transformations [18], although this counterexample does not satisfy the principle of reversibility—that any two pure states are linked by a reversible transformation—even on the level of the local state spaces.

A natural next step would be to investigate the maximal tensor products of models whose local state spaces are given by polygons [20] in order to determine whether the corresponding set of reversible dynamics is trivial. These models bear similarity to Boxworld in that their effect cones have a finite number of extreme rays, and that they are locally reversible; however, there are more linear relations between the extreme rays of polygon systems than there are in Boxworld. The similarity with Boxworld is even greater for polygons with an even number of sides, since in this case the set of extreme rays is identical with the set of fiducial effects. The techniques developed in this paper may be useful in characterizing the set of reversible transformations in these systems.

At present the authors do not know of any theory which satisfies reversibility and is not ‘embeddable’ within quantum theory. An interesting open conjecture is that any theory in the general probabilistic framework which satisfies local tomography and reversibility can be embedded within quantum theory. If this is the case, it would imply that quantum theory is maximal in the set of reversible theories, and that reversibility is sufficient as a ‘physical axiom’ from which to derive quantum theory. This would strongly suggest that the reversibility of physical laws lies at the heart of the obscure mathematical structure of quantum theory. Alternatively, any counterexample to this conjecture would be a fascinating foil theory with which to compare quantum theory, and perhaps to move beyond it.

Acknowledgments

AJS thanks the Royal Society for their support. SWA is funded by an EPSRC grant.

Appendix A

Lemma 1. Let \( E = \sum \alpha e_\alpha \) be an i-sub-unit effect. Then each \( e_\alpha \) differs from \( E \) only in component \( i \). Moreover, the set of ith components \( \{ e^{(i)}_\alpha \} \) forms a fiducial measurement on system \( i \).

Proof. We will prove the lemma by contradiction. Suppose first that \( e^{(i)}_\alpha \neq E^{(i)} \) for some \( \alpha ' \) and some \( j \neq i \). Let \( E^{(j)} = X^{(j)}_{\alpha ' (j)} \) and \( e^{(j)}_\alpha = X^{(j)}_{\alpha (j)} \). Either \( x_j \neq x_j ' \), or \( x_j = x_j ' \) but \( a_j \neq a_j ' \), so we can construct a pure product state \( s^{(j)} \) on system \( j \) such that \( s^{(j)}(j) \neq a_j \) and \( s^{(j)}(j) = a_j ' \), i.e. which hits \( e_\alpha \) but not \( E \). Then for any pure product state \( s \) whose \( j \)th component is \( s^{(j)} \), \( \langle E, s \rangle = 0 \) but \( \langle e_\alpha, s \rangle = 1 \), contradicting the fact that \( \langle e_\alpha, s \rangle \leq \langle E, s \rangle \). \( \{ e^{(i)}_\alpha \} \) is a set of fiducial effects satisfying \( \sum e^{(i)}_\alpha = \mathcal{U}^{(i)} \), hence any pure state \( s^{(i)} \) on system \( i \) must hit exactly one of the \( e^{(i)}_\alpha \). If any two of the \( e^{(i)}_\alpha \) are effects corresponding to different measurements, then there is a pure state \( s^{(i)} \) which hits both of them. Hence the effects all belong to the same fiducial measurement \( x \); if \( \{ e^{(i)}_\alpha \} \) is not the full set of outcomes of measurement \( x \), then there is a pure state \( s^{(i)} \) which hits none of them. It follows that \( \{ e^{(i)}_\alpha \} \) forms a fiducial measurement on system \( i \).

\[ \text{\[10\] i.e. For which there is a map from all states, measurements and transformations in the theory to quantum states, measurements and transformations, which recovers all predictions of the theory.} \]
Appendix B

Lemma 2. For $r \leq K_1^{(1)}$ suppose that $\{e_a\}_{a=1}^r$ does not cover any sub-unit effects. Then for any fiducial effect $f \not\in \{e_a\}$, there is a pure product state which hits $f$ but none of the $e_a$.

Proof. Let $f = X_{a_1}^{(1)} \otimes \cdots \otimes X_{a_N}^{(N)}$. We proceed by induction on the number of systems $N$. When $N=1$ set $s_{1} = e_{a_{1}}$ to ensure that $s$ hits $X_{a_{1}}^{(1)}$. The conditions imply that no partial sum of $\{e_a\}$ equals the unit effect, hence for each other choice of measurement $x' \neq x$, it must be possible to choose $s_x$ such that $X_{a_{1}/x'} \not\in \{e_a\}$. By construction $s$ hits $X_{a_{1}}^{(1)}$ but none of the $e_a$.

When $N > 1$, note that for any fiducial effect $g$ on system 1, the set $\{e_a^{[2, \ldots, N]} : e_a^{(1)} = g\}$ is a decomposition of some effect on the remaining $N-1$ systems with at most $K_1^{(1)} \leq K_1^{(2)}$ elements. This decomposition satisfies the conditions of the Lemma, hence by induction there exists a pure product state $s^{(2)} \otimes \cdots \otimes s^{(N)}$ which hits $f^{[2 \ldots N]}$ but none of the set $\{e_a^{[2, \ldots, N]} : e_a^{(1)} = g\}$.

Again, it is necessary to set $s^{(1)} = a_{1}$. Consider the set $\{e_a^{(1)}\}$, and the outcomes for measurements other than $x_1$ on system 1. One of two cases must occur:

(a) The set $\{e_a^{(1)}\}$ fills none of the other measurements, i.e. for every $x' \neq x_1$, there is an $a_{x'}$ such that $X_{a_{x'}}^{(1)} \not\in \{e_a^{(1)}\}$. For each such $x' \setminus x_1$, set $s^{(1)} = a_{x'}$ so that $s$ can hit $e_{a_{x'}}$ only if $e_{a_{x'}}^{(1)} = f^{(1)}$. However, using the inductive hypothesis, there exists a pure product state $s^{(2)} \otimes \cdots \otimes s^{(N)}$ which hits $f^{[2 \ldots N]}$ but none of the sets $\{e_a^{[2, \ldots, N]} : e_a^{(1)} = f^{(1)}\}$.

(b) There exists a measurement $x' \neq x_1$ on system 1 with $K_1^{(1)} = r$ which is filled by the set $\{e_a^{(1)}\}$, i.e. (after reordering) $e_a^{(1)} = X_{a_{x'}}^{(1)}$ for $1 \leq a \leq r$. $\{e_a\}$ covers no sub-unit effects, so there must be some $a'$ and some system $i \neq 1$ such that $e^{(1)}_a \neq f^{(1)}$. Set $s^{(1)} = a'$ so that $s$ does not hit any $e_{a'}$ with $a \neq a'$; the remaining components of $s^{(1)}$ may be chosen arbitrarily. By the inductive hypothesis, there exists a pure product state $s^{(2)} \otimes \cdots \otimes s^{(N)}$ which hits $f^{[2 \ldots N]}$ but not the single effect $e_a^{[2, \ldots, N]}$.

In both cases, by construction $s = s^{(1)} \otimes \cdots \otimes s^{(N)}$ hits $f$ but none of the $e_a$. \qed

Appendix C

It will be convenient to let $S_{a,i}$ denote the set of sub-unit effects at system $i$, and for a subset of systems $\Omega \subseteq [N]$, let $S_{\Omega} = \bigcup_{i \in \Omega} S_{i}$. The proof of lemma 5 relies on some simple results about reversible transformations.

Lemma 6. Let $\Omega \subseteq [N]$ and suppose that $T^\dagger$ is an allowed reversible transformation which permutes the set $S_{\Omega}$. Then the transformations of two fiducial effects will be identical outside $\Omega$ if and only if the original effects were. i.e.:

$$e_1^\Omega = e_2^\Omega \iff (T^\dagger(e_1))^\Omega = (T^\dagger(e_2))^\Omega$$

where $\Omega = [N] \setminus \Omega$.

Proof. Suppose firstly that the fiducial effects $e_1$ and $e_2$ differ only in one component $i \in \Omega$. $e_1$ and $e_2$ belong to (possibly different) decompositions of a unique sub-unit effect $E \in S_{i}$. By assumption $T'(E)$ is an $i'$-sub-unit effect for some $i' \in \Omega$; $T'(e_1)$ and $T'(e_2)$ belong to decompositions of $T'(E)$, hence by lemma 1 can only differ in component $i'$. 11
Suppose now that $e_1$, $e_2$ satisfy $e_1^{(k)} = e_2^{(k)}$ for all $k \notin \Omega$, but that they differ in any number of components belonging to $\Omega$. Then it is possible to move from $e_1$ to $e_2$ by changing one component at a time (each component belonging to $\Omega$). At each step, $T'$ maps the corresponding pair of effects to a pair which differ only in components belonging to $\Omega$. Hence $T' \{ e_i \} = T' \{ e_j \}$ for all $k \notin \Omega$.

To prove the converse direction, note that if $T'$ is an allowed reversible transformation which permutes the set $\mathcal{S}_T$, then so is $(T')^{-1}$.

Lemma 7. Suppose that $E = \sum_{n=1}^r s_n$ is a sub-unit effect, and that $\{ T' \{ e_n \} \}$ covers a sub-unit effect $F$. Then $T' \{ E \} = F$.

**Proof.** Without loss of generality let $\sum_{n=1}^r T' \{ e_n \} = F$ for $s \leq r$, and let $\sum_{\beta=1}^r f_\beta$ be a distinct decomposition of $F$. Then $E$ covers the multiform effect $(T')^{-1}(F) = \sum_{\alpha=1}^r e_\alpha = \sum_{\beta=1}^r (T')^{-1}(f_\beta)$. It follows from corollary 1 that $s = r$ and that $T'(E) = F$.

Lemma 8. Suppose that $\{ e_n \}$ does not cover any sub-unit effects, but that there exists some system $i$ for which $\sum_{n=1}^r e_n^{(i)} = \mathcal{U}^{(i)}$. Then for any fiducial effect $f \notin \{ e_n \}$, there exists a pure product state which hits $f$ but none of the $e_n$.

**Proof.** Let $f^{(i)} = X_{dl}^{(i)}$ and $\Omega_i = [N]\{i\}$. Note that $\{ e_a^{(i)} \}$ is the complete set of outcomes for some fiducial measurement $x'$ on system $i$: without loss of generality, $e_a^{(i)} = X_{dl}^{(i)}$.

If $x' = x$, then $f^{(i)} = e_a^{(i)}$. Let $s^{(i)} = a$ and choose the remaining components of $f^{(i)}$ arbitrarily, so that $s^{(i)}$ hits $e_a^{(i)}$ but none of the other $e_a^{(i)}$. Note that $f^{(i)}$ and $e_a^{(i)}$ must be distinct fiducial effects, so by lemma 2 there exists a pure product state $s^{(i)_\Omega}$ which hits $f^{(i)}$ but not $e_a^{(i)}$.

If $x' \neq x$, then since $\sum_{\alpha} e_\alpha$ is not a sub-unit effect, there exists $\alpha'$ and $i' \neq i$ such that $e_{\alpha'}^{(i')} \neq f^{(i')}$. Set $s_{\alpha'}^{(i')} = a$ and choose the remaining components of $f^{(i')}$ arbitrarily. By lemma 2 there is a pure product state $s^{(i')_\Omega}$ which hits $f^{(i')}$ but not the single fiducial effect $e_{\alpha'}^{(i')}$. In both cases, combining $s^{(i)}$ with $s^{(i')_\Omega}$ gives a pure product state $s$ which hits $f$ but none of the $e_n$.

We are now in a position to prove lemma 5 from the main text.

**Lemma 5.** Reversible Boxworld transformations map sub-unit effects to sub-unit effects, as long as none of the systems are classical.

**Proof.** We begin by considering the action of $T'$ on a 1-sub-unit effect $E$. $T'(E)$ is a multiform effect with a decomposition containing $K^{(1)}_i$ elements, hence by corollary 2 it is a $j$-sub-unit effect for some subsystem $j$ with $K^{(1)}_j = K^{(1)}_i$. By the same reasoning $T'$ permutes the set $\mathcal{S}_T$, where $\Omega = \{ j : K^{(1)}_j = K^{(1)}_i \}$.

We now show iteratively for each positive integer $r > K^{(1)}_i$ that $T'$ permutes the set $\mathcal{S}_{T_\Omega}$, where $\Omega = \{ i : K^{(1)}_i = r \}$. Let $i \in \Omega$, let $\sum_{\alpha=1}^r e_\alpha = \sum_{\beta=1}^r e_\beta$ be distinct decompositions of an $i$-sub-unit effect $E$ and assume that $T'$ permutes the set $S_{T_\Omega}$, where $\Omega = \{ j : K^{(1)}_j < r \}$. Note that $T'(E)$ is also multiform, since $T'(E) = \sum_{\alpha=1}^r T'(e_\alpha) = \sum_{\beta=1}^r (T')^{(i)}(e_\beta)$. Write $f_\alpha = T'(e_\alpha)$ and $g = T'(e_\alpha)$, noting that $g \notin \{ f_\alpha \}$.

Assuming that $\{ f_\alpha \}$ does not cover a sub-unit effect, our aim is to construct a pure product state $s$ that hits $g$ but none of the $f_\alpha$, giving a contradiction. Hence $\{ T'(e_n) \}$ must cover a
sub-unit effect. It then follows from lemma 7 that $T'(E)$ is itself an $i'$-sub-unit effect for some $i' \in \Omega_i$. By continuing the iteration, we complete the proof of the lemma.

To obtain the desired contradiction, assume that $\{|f_\alpha\}$ does not cover a sub-unit effect. Let $\bar{\Omega} = [N] \setminus \Omega$ and consider the set $\{|f_\alpha^\Omega\}_{\alpha=1}^r$ (recall that for a fiducial effect $f$, $f^\Omega$ is the tensor product of all those components of $f$ belonging to $\bar{\Omega}$). Since $e^{(\alpha)}$ is distinct from $e^{(\alpha)}_{\bar{\Omega}}$ for all $\alpha$, and $i \in \bar{\Omega}$, we have that $e^{(\alpha)} \notin \{|e^{(\alpha)}_{\bar{\Omega}}\}$. It follows from lemma 6 that $g^\bar{\Omega} \notin \{|f_\alpha^\bar{\Omega}\}$.

Suppose that there exists a subsystem $i' \in \bar{\Omega}$ such that $\{|f_{\alpha i'_1}^{(\alpha)}\}_{\alpha=1}^r$ covers the local unit effect $\mathcal{U}(\alpha)$. Recall that the fiducial effect decompositions of $\mathcal{U}(\alpha)$ are obtained by fixing a fiducial measurement on subsystem $i'$ and taking the set of local fiducial effect vectors which correspond to an outcome of that measurement. Since all fiducial measurements on subsystem $i'$ have at least $r$ outcomes, it follows that $\sum_{\alpha} f_{\alpha i'_1}^{(\alpha)} = \mathcal{U}(\alpha)$, and by lemma 8 there exists a state which hits $g$ but none of the $f_\alpha$. Suppose instead that there is no subsystem $i' \in \bar{\Omega}$ for which $\{|f_{\alpha i'_1}^{(\alpha)}\}_{\alpha=1}^r$ covers $\mathcal{U}(\alpha)$. Then the set $\{|f_\alpha^\bar{\Omega}\}$—considered as a collection of fiducial effects over the maximal tensor product of all subsystems belonging to $\bar{\Omega}$—does not cover any sub-unit effects. By lemma 2 applied to the subsystems belonging to $\bar{\Omega}$, there exists a pure product state $s^\bar{\Omega}$ which hits $g^\bar{\Omega}$ but none of the $f_\alpha^\bar{\Omega}$. Combining $s^\bar{\Omega}$ with any pure state $s^\Omega$ which hits $g^\Omega$ gives a pure product state $s$ which hits $g$ but none of the $f_\alpha$. ⊓⊔

References

[1] Barrett J 2007 Information processing in generalized probabilistic theories Phys. Rev. A 75 032304
[2] Hardy L 2001 Quantum theory from five reasonable axioms arXiv:quant-ph/0101012
[3] Chiribella G, D’Ariano G M and Perinotti P 2010 Probabilistic theories with purification Phys. Rev. A 81 062348
[4] Chiribella G, D’Ariano G M and Perinotti P 2011 Informational derivation of quantum theory Phys. Rev. A 84 012311
[5] Short A and Barrett J 2010 Strong nonlocality: a trade-off between states and measurements New J. Phys. 12 033034
[6] Bell J S 1965 On the Einstein–Podolsky–Rosen paradox Physics 1 195–200
[7] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom Found. Phys. 24 379–85
[8] van Dam W 2005 Implausible consequences of superstrong nonlocality arXiv:quant-ph/0501159
[9] Buhrman H, Christandl M, Unger F, Wehner S and Winter A 2006 Implications of superstrong nonlocality for cryptography Proc. R. Soc. A 462 1919
[10] de la Torre G, Masanes L, Short A J and Müller M P 2012 Deriving quantum theory from its local structure and reversibility Phys. Rev. Lett. 109 090403
[11] Müller M P and Ududec C 2012 The structure of reversible computation determines the self-duality of quantum theory Phys. Rev. Lett. 108 130401
[12] Chiribella G, D’Ariano G M and Perinotti P 2012 Quantum theory, namely the pure and reversible theory of information Entropy 14 1877
[13] Masanes L and Müller M P 2011 A derivation of quantum theory from physical requirements New J. Phys. 13 063001
[14] Gross D, Müller M, Colbeck R and Dahlsten O C O 2010 All reversible dynamics in maximally non-local theories are trivial Phys. Rev. Lett. 104 080402
[15] Perez-Garcia D, Augusiak R, Masanes Ll and Müller M P 2011 Entanglement and the three-dimensionality of the bloch ball arXiv:quant-ph/1111.4060
[16] Dakić B and Brukner C 2009 Quantum theory and beyond: is entanglement special? arXiv:quant-ph/0911.0695
[17] Hardy L 2011 Reformulating and reconstructing quantum theory arXiv:quant-ph/1104.2066
[18] Al-Safi S W 2014 Quantum theory from the perspective of general probabilistic theories PhD thesis (in preparation)
[19] Leifer M, Wilce A, Barnum H and Barrett J 2006 Cloning and broadcasting in generic probabilistic theories arXiv:quant-ph/0611295
[20] Janotta P, Gogolin C, Barrett J and Brunner N 2011 Limits on nonlocal correlations from the structure of the local state space New J. Phys. 13 063024