

A counterexample to a positive entropy skew product generalization of the Pinsker conjecture

JONATHAN KING

Department of Mathematics, SUNY at Albany, Albany, NY 12222, USA

(Received 11 June 1984 and revised 19 November 1984)

Abstract. The class of \mathcal{K} -automorphisms is not contained in a certain class of skew products over a Bernoulli base. The non-identity fibre transformation in the skew is allowed to have positive or even infinite entropy. A difficulty presented by positive entropy is handled via an apparently new property of independent processes (lemma 7.24).

1. Introduction

(1.1) Statement of problem.

Definition. By a *simple independent skew* we shall mean the following: On the base we have a Bernoulli transformation \mathcal{B} acting on the space $\Omega_{\mathcal{B}}$; two subsets $\mathbf{H}, \mathbf{T} \subset \Omega_{\mathcal{B}}$ (Heads and Tails) whose disjoint union is $\Omega_{\mathcal{B}}$, such that the two set partition $\langle \mathbf{H}, \mathbf{T} \rangle$ is independent and generates under \mathcal{B} . Also, we have an ergodic process \mathcal{Z} on the space $\Omega_{\mathcal{Z}}$. For fibre transformations we have the identity over \mathbf{T} and the transformation \mathcal{Z} over \mathbf{H} . Our skew product \mathcal{S} acts on the product space $\Lambda = \Omega_{\mathcal{B}} \times \Omega_{\mathcal{Z}}$ as follows: Any point $\lambda = \langle \beta, \zeta \rangle \in \Lambda$ gets mapped by \mathcal{S} to

$$\mathcal{S}\langle \beta, \zeta \rangle = \begin{cases} \langle \mathcal{B}\beta, \mathcal{Z}\zeta \rangle & \text{if } \beta \in \mathbf{H}; \\ \langle \mathcal{B}\beta, \zeta \rangle & \text{if } \beta \in \mathbf{T}. \end{cases}$$

A shorthand for this skew product will be $\mathcal{B} \otimes (\mathcal{Z}/\mathbf{H}, \text{Id}/\mathbf{T})$. Pictorially more suggestive than $\langle \beta, \zeta \rangle$, we shall use $\langle \beta^{\zeta} \rangle$ to denote a point in the space of the skew product. We shall refer to ζ and β as the \mathcal{Z} -component, respectively \mathcal{B} -component, of λ , where $\lambda = \langle \beta^{\zeta} \rangle$.

The goal of this paper is to produce a \mathcal{K} -automorphism \mathcal{T} which is not isomorphic to any simple independent skew having $0 < \mu(\mathbf{H}) < 1$.

(1.2) *Raison d'être.* Suppose we have a transformation \mathcal{T} represented as a skew product of a family of fibre transformations $\{\mathcal{Z}_{\beta} : \beta \in \Omega_{\mathcal{B}}\}$ over a base transformation \mathcal{B} . How much can we tell about the entropy of \mathcal{T} by knowing the entropy $E(\mathcal{B})$ as well as all the entropies $\{E(\mathcal{Z}_{\beta})\}_{\beta}$? Certainly $E(\mathcal{T}) \geq E(\mathcal{B})$ because \mathcal{B} is a factor of \mathcal{T} . But how much entropy does the family $\{\mathcal{Z}_{\beta}\}_{\beta}$ add? It turns out there is no particular connection between the set of numbers $\{E(\mathcal{Z}_{\beta})\}_{\beta}$ and the number $E(\mathcal{T}) - E(\mathcal{B})$. For example, it is possible to have $E(\mathcal{Z}_{\beta}) = \infty$ for every β and yet have

$E(\mathcal{T}) = E(\mathcal{B}) = 0$. In the other direction, we can have each $E(\mathcal{X}_\beta) = 0$ for every β and yet have $E(\mathcal{T}) = \infty$ and $E(\mathcal{B}) = 0$.

The question then arises: For an arbitrary \mathcal{H} -automorphism \mathcal{T} does there exist a skew representation as above where \mathcal{B} is a Bernoulli factor of maximal entropy and each \mathcal{X}_β has zero entropy? In this generality the question may be difficult (particularly if true). However, this problem becomes tractable if we place restrictive conditions on the type of skew we allow. The most basic non-trivial skew product would have just two kinds of fibre transformations, of which one was the identity. The simple independent skews naturally suggest themselves.

What are some of the properties that make simple independent skews easier to study? One useful property is that if \mathbf{P} is a generating partition for \mathcal{X} , then the product partition $\mathbf{P}' = \langle \mathbf{H}, \mathbf{T} \rangle \times \mathbf{P}$ is a generating partition for \mathcal{S} . Why? Let us take a look at the \mathcal{S} - \mathbf{P}' -name of a point $\lambda = \langle \zeta, \beta \rangle$. The upper component, ζ , is a doubly infinite \mathcal{X} - \mathbf{P} -name. The lower component, β , is a \mathcal{B} - $\langle \mathbf{H}, \mathbf{T} \rangle$ -name. λ 's i th letter, λ_i , is a pair of letters ζ_{β_i} . As an example, suppose $\beta = \cdots \beta_{-1} \tilde{\beta}_0 \beta_1 \beta_2 \beta_3 \tilde{\beta}_4 \tilde{\beta}_5 \beta_6 \cdots$ where we place an arrow over a β_i to indicate that it is 'H'. Then the name λ will look like the sequence of letter-pairs in the bottom row, shown below:

$$\begin{array}{cccccccccc} \cdots & \lambda_{-1} & \lambda_0 & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 & \lambda_5 & \lambda_6 & \cdots \\ \cdots & \zeta_0 & \zeta_0 & \zeta_1 & \zeta_1 & \zeta_1 & \zeta_1 & \zeta_2 & \zeta_3 & \cdots \\ \cdots & \beta_{-1} & \tilde{\beta}_0 & \beta_1 & \beta_2 & \beta_3 & \tilde{\beta}_4 & \tilde{\beta}_5 & \beta_6 & \cdots \end{array}$$

Hence, if we have two distinct points $\langle \zeta, \beta \rangle, \langle \zeta', \beta' \rangle \in \Lambda$ then either $\zeta \neq \zeta'$ or $\beta \neq \beta'$ and consequently their \mathbf{P}' -names are different. We see that the 'product letter' λ_i is the pair of letters ζ_{β_i} where $w = (\text{the number of Heads in } \beta_0, \dots, \beta_{i-1})$. Another way to say this is that

$$\mathcal{S}^i \lambda = \mathcal{S}^i \langle \zeta, \beta \rangle = \left\langle \mathcal{X}^w \zeta, \mathcal{B}^i \beta \right\rangle.$$

This motivates the following.

Definition. We define the function $\text{Weight}(\text{finite string of Heads and Tails})$ to be the number of Heads in that finite string. Let $\beta|_i^j$ denote the substring $\beta_i \beta_{i+1} \cdots \beta_{j-1}$ of β . By extension, use $|_i^j$ to mean the 'half open' interval of integers $[i, j)$. Also, $\beta|_i$ abbreviates $\beta|_i^{i+1}$. Finally, let $\mathcal{B}|_a^b$ denote the set of finite names $\{\beta|_a^b : \beta \in \Omega_{\mathcal{B}}\}$. Viewed as a set of names, $\mathcal{B}|_a^b$ is no different from the set $\mathcal{B}|_0^{b-a}$ - but sometimes it is convenient to emphasize where it is in \mathcal{B} -names that our attention is focused. Viewed as a union of cylinder sets, $\mathcal{B}|_0^{b-a}$ is the image of $\mathcal{B}|_a^b$ under the transformation \mathcal{B}^a .

Here is how one can think of the action of \mathcal{S} on $\langle \zeta, \beta \rangle$: The index i runs along β at a constant rate. The sequence of Heads and Tails in β tells us how to 'read' ζ ; the ζ -index w moves along ζ in fits and starts, as i moves, always keeping $w = \text{Weight}(\beta|_0^i)$. Although w 's small-scale behaviour is quite random, the strong law of large numbers assures us that w 's average speed is $\mu(\mathbf{H})$; when i is large, w will be within a small percentage of $i \cdot \mu(\mathbf{H})$.

The above considerations give rise to a second nice property of simple independent skewes, namely that we can calculate the entropy of the skew product explicitly from those of the base and the fibres. The relation is

$$E(\mathcal{S}) = E(\mathcal{B}) + E(\mathcal{Z})\mu(\mathbf{H}).$$

We will not use this fact so we give the proof modulo the details: Let E and E' denote the respective entropies of \mathcal{B} and \mathcal{Z} . So, for a given i and w , we expect to see about $2^{E \cdot i}$ many \mathcal{B} - i -names and $2^{E' \cdot w}$ \mathcal{Z} - w -names, respectively, of significant measure. Now fix a large i and let $w = \text{Weight}(\beta|_0^i)$. For a typical $\langle \beta \rangle$ we know that $w \approx i \cdot \mu(\mathbf{H})$. And the number of significant \mathcal{S} - i -names is the product of the number of significant \mathcal{B} - i -names with the number of significant \mathcal{Z} - w -names. This is about $2^{Ei} \cdot 2^{E'w} \approx 2^{(E+E'\mu(\mathbf{H}))i}$ and our conclusion follows by an appeal to the Shannon-MacMillan-Breiman Theorem.

Is there still content? Have we trivialized the problem by our restriction – is the class of simple independent skewes sufficiently rich to be interesting?

Yes. With a bit of argument, one can show (see [4]) that any simple independent skew with an ergodic fibre transformation \mathcal{Z} is automatically a Kolmogorov automorphism. Moreover, the class contains ‘non-trivial’ \mathcal{K} -automorphisms, that is to say, transformations not isomorphic to a Bernoulli. In fact, Jack Feldman used a simple independent skew to construct a transformation which is not even Kakutani equivalent to a Bernoulli (see [7]). Other interesting transformations have been found in this class (see [2] and [10]). So it is not unnatural to wonder whether all \mathcal{K} -automorphisms could be built with such a skew.

The second motivation comes from trying to resuscitate the Pinsker conjecture. Pinsker conjectured that every transformation could be written as the direct product of a Bernoulli cross a zero entropy transformation; Don Ornstein built a counterexample of this in [5]. But, conceivably, a variant of Pinsker’s conjecture could have been rescued by generalizing the direct product (the case where $\mu(\mathbf{H}) = 1$) to simple independent skew (the case when $0 < \mu(\mathbf{H}) < 1$). Our result says no. In fact, by a bit of serendipity, we were able to remove the restriction that $E(\mathcal{Z}) = 0$. As a consequence, our result bears some resemblance to, and could conceivably shed some light on, the ‘Weak Pinsker Conjecture’ (which asserts that every transformation can be written as a direct product of the form Bernoulli \times \mathcal{Z} for transformations \mathcal{Z} of arbitrarily small entropy).

(1.3) *Notation and conventions.* The measure on any of our spaces we denote by $\mu(\cdot)$. Definitions and assertions are intended to hold ‘almost everywhere’. As is customary, the ‘a.e.’ will usually be tacit.

$a \triangleq b$ means the expressions a and b are equal by definition. In contrast, I use $a \trianglelefteq b$ or $b \trianglelefteq a$ to mean: the expression b defines the symbol a . This can conveniently be extended so that a statement like $(1 - \alpha)/2 \trianglelefteq 7$ means ‘define the symbol α so that $(1 - \alpha)/2 = 7$ holds’.

The expression $\forall_{\text{large}} n$ means ‘ $\exists N$ such that $\forall n > N$ ’. I use a question mark to mean some number for which I do not need or want a name. As an example, one could say ‘Consider a finite set of integers of the form $\{1, 2, \dots, ?\}$ ’.

A common type of argument, which we will just call ‘Chebyshev’, is the following: If most pairs $\langle x, y \rangle$ have a certain property, then for most x , most y are such that $\langle x, y \rangle$ has the property. Of course the first ‘most’ needs to be of the form $1 - \delta^2$ and the next two ‘most’s are each $1 - \delta$.

This argument appears frequently and I will usually omit the details. In the same spirit, statements of theorems will sometimes contain phrases like

\forall ‘most’, *most x have the property that . . .*

This convention simplifies and makes more intuitive the exposition of trivial implications such as ‘*Most x have property 1 and most x have property 2 so most x have properties 1 and 2*’, where the three ‘most’s are of the form $1 - \delta$, $1 - \delta$, and $1 - 2\delta$, but where this specific relationship is irrelevant because δ may be made arbitrarily small.

Some definitions later on start off with the words, ‘for any fixed, implicit, $n \dots$ ’. The n will indicate the n th stage of the construction that builds our counterexample. Nearly every symbol appearing will depend on this n . It seems to me to be overly ‘noisy’ notation to have a subscript n attached to each term; by ‘fixed, implicit’ I mean ‘not explicitly appearing in the notation’.

The convention on the various appendages one can put on a name is: τ, β, ζ are doubly infinite names i.e. points, in their respective spaces $X, \Omega_{\mathcal{B}}$, and $\Omega_{\mathcal{X}}$. $\bar{\tau}, \bar{\beta}$ are finite strings of length $h(n)$ corresponding to n -blocks. $\check{\tau}, \check{\beta}$ are short substrings of n -blocks.

The expression ‘ ε -percent’ means, ‘a fraction thereof, of size ε ’, where ε is some number between zero and one. For example, it would be valid to say ‘the set $\{2, 5, 6\}$ is $\frac{1}{3}$ -percent of the set $\{1, 2, \dots, 9\}$ ’.

The symbol \square ends a proof.

2. A brief overview

The overview presented here is imprecise and at times refers to terms not yet defined. Nonetheless, the hope is that the reader will gain a rough idea of the structure of the proof. Then, when proceeding through the details of later sections, one will be able to refer back to this overview in order to see how these details fit into the overall picture.

Our proof splits naturally into three phases: assembly, oracles, and contradiction from random overlap.

(2.1) *Assembly.* First, we shall construct a transformation \mathcal{T} and generating partition \mathbf{A} on a probability space X via cutting and stacking. Any point $x \in X$ has a name, τ , of its orbit relative to \mathbf{A} . We shall call this the *base spelling* of the point x . But this base spelling is too detailed to be useful; we miss the forest for the trees. Rather, we shall extract from τ a cruder spelling called its *n -spelling*.

Stage n of the cutting and stacking will specify how to section τ into special substrings each having the same length $h(n)$. In fact, these ‘special’ strings will simply be the columns of the n -gadget. We take this collection of special substrings and split it into subcollections $\xi_1, \xi_2, \dots, \xi_r$ which we call the (pseudo)letters, or n -letters, of the (pseudo) alphabet ${}_n\mathbf{A}$. We denote the number of n -letters by $|{}_n\mathbf{A}|$.

To obtain the n -spelling of τ , we scan along τ and replace each occurrence of a special substring by the ξ_i , which names the subcollection of which the special substring is a member. This gives a spelling like:

$$\dots, \xi_{i(-1)}, \xi_{i(0)}, \xi_{i(1)}, \xi_{i(2)}, \dots$$

While we have thrown away information in passing from τ to its n -spelling, we have not lost all control. For it will be the case that members of distinct n -letters look starkly different, even over small substrings. In fact, if $\check{\tau}$ is any not-too-small substring of a member $\bar{\tau} \triangleq \tau|_a^{a+h(n)}$ of an n -letter ξ_i , then, even though we cannot reconstruct the string $\bar{\tau}$ from its small substring $\check{\tau}$, we can determine the n -letter ξ_i that $\bar{\tau}$ belongs to.

To start the next stage of the argument we postulate, for the sake of contradiction, that we have an isomorphism φ from some simple independent skew $\mathcal{B} \otimes \langle \mathcal{X}/\mathbf{H}, \text{Id}/\mathbf{T} \rangle$ to our transformation \mathcal{T} . We use φ^{-1} to lift the concept of n -spelling to the skew product space and can now speak of the n -spelling of a point $\langle \beta \rangle$. We denote the k th letter of this n -spelling by $({}_n A \langle \beta \rangle)_k$.

(2.2) *Oracles.* It follows from the fact that members of different n -letters look starkly different that the n -spelling of a point $\langle \beta \rangle$ is approximately independent of β and can be determined, in a rough sense, by examining just the upper component ζ . The tool we introduce is the concept of an *oracle*. An oracle is a substring of ζ , quite short compared to the length $h(n)$ of an n -letter, with predictive power. If $\zeta|_0^{\varepsilon \cdot h(n)}$ is an oracle, then we can make a pretty good guess as to which n -letter the point $\langle \beta \rangle$ is in at time zero – ignorant though we are of β . What we will *not* know is *which* member of the n -letter it is in – *that* depends very much on β .

In § 6 we show that oracles are a common occurrence. If we scan a typical \mathcal{X} -name $\zeta|_0^\infty = \zeta_0 \zeta_1 \dots \zeta_w \dots$ as the index w runs from 0 towards ∞ , we shall find that for a high density of values w , ζ_w is the first letter of an oracle.

At this juncture one may be tempted to remark: even if $\zeta|_0^\infty$ hits oracles frequently, that gives us but a nebulous view of the n -spelling of $\langle \beta \rangle$. Not knowing β , we will not know its pattern for reading ζ . We may see an oracle commencing at, say, ζ_{17} which suggests strongly that a particular n -letter ξ appears somewhere in the n -spelling of $\langle \beta \rangle$ – but its exact location is unknown.

So, rather than fight it, we take advantage of the fact that the pattern in which the n -letters of $\langle \beta \rangle$ sit on ζ is, in the long run, random.

(2.3) *Contradiction from random overlap.* Oracles are very short creatures compared with $h(n)$, the length of an n -letter. Indeed, oracles have length $\varepsilon \cdot h(n)$ for as Lilliputian an ε as we are willing to pay for by making n large. But as short as they are, oracles force a certain long-distance regularity.

Suppose $\langle \beta \rangle|_0^{h(n)}$ is a member of an n -letter ξ . As the index w runs along ζ from 0 to the number $\text{Weight}(\beta|_0^{h(n)})$, it is likely to find many ζ_w which commence oracles. It is not, then, too improbable to find an oracle near each end of, and in, the index interval $[0, \text{Weight}(\beta|_0^{h(n)})]$. These two oracles are quite far apart compared to their lengths and yet they almost certainly make identical predictions.

Now pick two typical points in the product space with the same \mathcal{L} -component – say $\langle \xi_\beta \rangle$ and $\langle \xi_\rho \rangle$. β and ρ are two randomly chosen Bernoulli names and consequently each is ignorant of how the other splits their common \mathcal{L} -component ζ into n -letters. What typically happens, when we lay both sets of hashmarks down on ζ ? For convenience, we simplify the notation by assuming that time zero of the point $\langle \xi_\beta \rangle$ happens to be the first coordinate of an n -block. So the n -spelling of $\langle \xi_\beta \rangle$ puts hashmarks in β at locations

$$i = 0, h(n), 2h(n), \dots, kh(n), \dots$$

Where do these hashmarks fall in ζ ? At $w_0, w_1, \dots, w_k, \dots$ where $w_k \triangleq \text{Weight}(\beta|_0^{kh(n)})$. Hence, an oracle commencing at ζ_w , where $w \in]w_k, w_{k+1}[$, makes a prediction about the k th letter of the n -spelling of $\langle \xi_\beta \rangle$.

The strong law of large numbers tells us that the $\{w_k\}_k$, which are the hashmarks that ζ gets from β , are typically about distance $h(n) \cdot \mu(\mathbf{H})$ apart, one from the next. Similarly, ζ inherits hashmarks from ρ – denote them $\{W_l\}_{l=0}^\infty$, also typically separated by roughly $h(n)\mu(\mathbf{H})$. However, neither set of hashmarks can have any knowledge of the other’s spacing, in the long run. Consequently, as we scan along $\zeta|_0^\infty$ we often find indices

$$w_k < W_l < w_{k+1} < W_{l+1} < w_{k+2}$$

with each pair separated by a distance large against $\varepsilon h(n)$. Then $({}_n\mathbf{A}\langle \xi_\beta \rangle)|_k$ and $({}_n\mathbf{A}\langle \xi_\rho \rangle)|_l$ overlap on $[W_l, w_{k+1})$ and $({}_n\mathbf{A}\langle \xi_\beta \rangle)|_{k+1}$ and $({}_n\mathbf{A}\langle \xi_\rho \rangle)|_{l+1}$ overlap on $[w_{k+1}, W_{l+1})$. If we have made the density of oracles high, by having chosen n large, then it is likely that we will find an oracle in each overlap which predicts correctly for each of the $\langle \xi_\beta \rangle$ and $\langle \xi_\rho \rangle$ n -letters that overlap there. Consequently,

$$({}_n\mathbf{A}\langle \xi_\beta \rangle)|_k = ({}_n\mathbf{A}\langle \xi_\rho \rangle)|_l = ({}_n\mathbf{A}\langle \xi_\beta \rangle)|_{k+1}.$$

We are forced to conclude that two consecutive n -letters of $\langle \xi_\beta \rangle$ are equal.

This is a typical event and must happen rather frequently along $({}_n\mathbf{A}\langle \xi_\beta \rangle)|_0^\infty$. But no; for at assemblytime we can insure that on any typical name, the n -letters are laid down in a fairly independent way. Since $|{}_n\mathbf{A}|$ goes to infinity as $n \rightarrow \infty$, we can have chosen n large enough that the probability of consecutive duplicate letters in $({}_n\mathbf{A}\langle \xi_\beta \rangle)|_0^\infty$ is rather small.

3. Assembly

This section has three parts. The first describes a technique for building transformations. In the second part, we use this technique to build an ergodic transformation \mathcal{F} having a certain property with respect to the \bar{d} -metric. In the last part, we modify \mathcal{F} and incorporate two additional properties. The resulting \mathcal{H} -automorphism, \mathcal{T} , shall be our counterexample transformation. This transformation \mathcal{T} is modelled on an unpublished example of Don Ornstein and Benjamin Weiss, which, in its turn, is based on the non-Loosely Bernoulli example of Jack Feldman. The Ornstein–Weiss example is a transformation which is not Kakutani equivalent to any direct product of a Bernoulli with a zero entropy transformation.

(3.1) *Cutting and stacking.* For a detailed exposition of cutting and stacking see [3, p. 91]. We give a brief generic description of the kind of stacking we need.

The idea is to define, in stages, a transformation \mathcal{T} and generating partition (alphabet) $\mathbf{A} = \{a, b, \dots, ?\}$ on the half-open unit interval $[0, 1)$. At stage n we will have a number ${}_n a$, so that $0 \leq {}_n a \leq {}_{n+1} a \leq 1$ and ${}_n a \rightarrow 1$. Our partition \mathbf{A} and \mathcal{T} (or \mathcal{T}^{-1}) will be defined on every point of $[0, {}_n a)$ and be as yet undefined on $[{}_n a, 1)$. Specifically, $[0, {}_n a)$ will have been cut up into some number, $h(n)$, of equal length half-open intervals $I_1, \dots, I_{h(n)}$ called *levels*. We view these levels as stacked one above the other with I_{i+1} above I_i and call this stack the *n-stack*. $h(n)$ is the *height* of the *n-stack* and I_1 is the *base* of the stack. \mathcal{T} is taken to map each point of the stack upward to the point above it in the next level. This defines \mathcal{T} on $I_1, \dots, I_{h(n)-1}$ and \mathcal{T}^{-1} on $I_2, \dots, I_{h(n)}$. Moreover, the partition \mathbf{A} will have been defined on $\bigcup_{i=1}^{h(n)} I_i$ in such a way that I_1 splits into a finite number of equal width half-open intervals S_1, \dots, S_r where each S_j is the set of all points in the base I_1 having some particular \mathcal{T} - \mathbf{A} - $h(n)$ -name. Each S_j is the base of an *n-column*: the set $\bigcup_{i=0}^{h(n)-1} \mathcal{T}^i S_j$. The *n-stack*, split into these columns by \mathbf{A} , is called the *n-gadget*.

How do we build the $(n+1)$ -gadget from the *n-gadget*? Pick integers L, L' , and M which, in general, will depend on n . Then subdivide all of the *n-columns* into thinner columns each having the same width w and so that, in total, there are $L \cdot M$ of these thin columns. List them in some order as $C_1, \dots, C_{L \cdot M}$. In order to extend the definition of \mathcal{T} and \mathbf{A} to more of our unit interval, we now pick some ${}_{n+1} a$, ${}_n a \leq {}_{n+1} a \leq 1$, and split the piece $[{}_n a, {}_{n+1} a)$ of our space into little half-open intervals $J_1, \dots, J_{L \cdot M}$ each, like the $\{C_i\}$, having width w . Extend \mathbf{A} to $[{}_n a, {}_{n+1} a)$ by assigning to each J_i a letter of \mathbf{A} . The $(n+1)$ -gadget will consist of M 'tall' columns manufactured as follows: Choose L of the thin columns $\{C_i\}$ and L' of the little intervals $\{J_i\}$. Treating the little intervals as columns of height 1, stack these $L+L'$ many columns one above the other, in some order, to form a tall column of height $L \cdot h(n) + L'$ and of width w . Thus $h(n+1) = Lh(n) + L'$.

The material $\bigcup_1^{L \cdot M} J_i$ is sometimes called 'filler' – the filler added at stage n .

(3.1.1) *Cutting and stacking viewed as concatenation of words.* The reader may verify that, in the limit as $n \rightarrow \infty$, we define a.e. on the space $[0, 1)$ a bi-measure preserving transformation \mathcal{T} and partition \mathbf{A} . Also, the cutting and stacking can be done to insure that \mathbf{A} is a generating partition and hence we can view the points in the space as doubly infinite \mathbf{A} -names. (There is the small technical point of insuring that if the \mathbf{A} - $h(n)$ -name of a point x , in the space, is the name of some *n-column*, then x is, in fact, in the base of the *n-gadget*. This can be done by insuring that there is some filler added at the *n*th stage and labelling this filler in some way with the number n . In this paper I have dispensed with the distraction of explicitly labelling the bases of my *n-gadgets*.) Thus, we see that our cutting and stacking is described by an alphabet \mathbf{A} , a sequence of numbers $h(1), h(2), \dots, h(n), \dots$, and a rule which specifies, given a collection of \mathbf{A} - $h(n)$ -words (the *n-columns*), how to concatenate copies of these words and copies of the letters of \mathbf{A} (the filler added at stage n), to form words of length $h(n+1)$.

We would like to know the converse: When does a rule for concatenating \mathbf{A} - $h(n)$ -words (usually called *n-blocks*) into \mathbf{A} - $h(n+1)$ -words (the $(n+1)$ -blocks) describe a valid cutting and stacking? Knowing this would allow us to build a

stationary stochastic process without explicitly mentioning the measure, that is, without having to specify explicitly the numbers w and ${}_n a$ at each stage.

In order to know that a rule describes a valid cutting and stacking, it is sufficient to check two conditions:

- (i) Each n -block appears with equal frequency in the $(n + 1)$ -gadget.

This assures that the $(n + 1)$ -blocks are consistent with saying that all the n -columns have equal width or, in other words, equal measure. This condition will be true trivially for each of the transformations we construct in the next section.

The above condition assures that the limit \mathcal{T} will be a bi-measure preserving map on some measure space – but this space may be infinite if we add too much filler at each stage n . To make the space finite, we must assure that the fraction of the $(n + 1)$ -gadget which is filler added at stage n , is a summable function of n :

- (ii) $\sum_{n=1}^{\infty} (\# \text{ of levels of filler added at stage } n)/h(n+1) < \infty$,

where the numerator means the number L' of the above. For the specific transformation \mathcal{T} constructed in the section to follow, this number L' will be made equal to $h(n)$. Thus it will be sufficient to check that $\sum_{n=1}^{\infty} h(n)/h(n+1) < \infty$.

(3.2) *The Feldman transformation.* In order to exhibit clearly the main idea without cluttering it up with secondary details, we first reiterate some of Jack Feldman’s construction of a (it turns out to be) zero entropy, ergodic transformation with a rather curious property with regard to the \bar{d} metric. We shall build this transformation \mathcal{F} and generating partition ${}_0\mathbf{A}$ via the foredescribed cutting and stacking method. It will require no filler.

We let ${}_n\mathbf{A}$ denote the set of columns of this n -gadget; so ${}_n\mathbf{A}$ is a set of words of length $h(n)$. Agree to call ${}_n\mathbf{A}$ the n th pseudo alphabet. Each string $\bar{\tau} \in {}_n\mathbf{A}$ is a pseudo letter. In general, by a pseudo alphabet \mathbf{A} , we mean a set of equal length words over some base alphabet. Let $\text{Len}(\mathbf{A})$ denote their common length; $|\mathbf{A}|$ is the pseudo alphabet’s size, that is, the number of pseudo letters.

(3.2.1) *Definition.* Let $\bar{d}(\cdot, \cdot)$ denote the usual Hamming metric on strings. ‘Sub-string’ has the usual meaning of a subsequence of consecutive letters of a given string. For any $\varepsilon \in (0, 1]$ define the function \bar{d}_ε on pairs of equal length words $\bar{\sigma}, \bar{\tau}$ as: $\bar{d}_\varepsilon(\bar{\sigma}, \bar{\tau})$ is the infimum of $\bar{d}(\check{\sigma}, \check{\tau})$ as $\check{\sigma}$ and $\check{\tau}$ range over all equal length substrings of $\bar{\sigma}$ and $\bar{\tau}$, respectively, of relative length at least ε -percent i.e. such that

$$\text{Len}(\check{\sigma}) = \text{Len}(\check{\tau}) \geq \varepsilon \cdot \text{Len}(\bar{\tau}).$$

For any pseudo alphabet \mathbf{A} , let $\bar{d}_\varepsilon(\mathbf{A})$ denote the number

$$\inf \{ \bar{d}_\varepsilon(\bar{\sigma}, \bar{\tau}) : \bar{\sigma}, \bar{\tau} \in \mathbf{A}; \bar{\sigma} \neq \bar{\tau} \}.$$

Our goal is now easy to state: Fixing some particular number $g \in (0, 1)$, we want our n -gadgets to be such that

$$(3.2.2) \quad \forall \varepsilon, \forall \text{large } n, \text{ we have that } \bar{d}_\varepsilon({}_n\mathbf{A}) > g.$$

(3.2.3) *Feldman spelling.* Given an alphabet $\mathbf{B} = \{a, b, \dots, z\}$, with $|\mathbf{B}| \cong l$, and a positive integer m , we can form m different sequences over \mathbf{B} , each of length l^m ,

which we call *Feldman spellings*. Denote the i th such by $\mathbf{B}_{i,m}$ as i ranges over $i = 1, 2, \dots, m$.

$$\mathbf{B}_{i,m} \triangleq \underbrace{\text{Cycle}_i \text{ Cycle}_i \cdots \text{Cycle}_i}_{l^{m-i} \text{ times}} \tag{1}$$

where

$$\text{Cycle}_i \triangleq \underbrace{\text{aaaa} \cdots \text{abbbb} \cdots \text{b}}_{l^{i-1} \text{ times}} \cdots \underbrace{\text{zzzz} \cdots \text{z}}_{l^{i-1} \text{ times}} \tag{1'}$$

Recall that there are l of these.

Evidently $\text{Len}(\text{Cycle}_i) = l^i$ and hence $\text{Len}(\mathbf{B}_{i,m}) = l^{m-i} \cdot l^i = l^m$. More generally, if we permit \mathbf{B} to be a pseudo alphabet then $\text{Len}(\mathbf{B}_{i,m}) = l^m \cdot \text{Len}(\mathbf{B})$.

Given an alphabet or pseudo alphabet \mathbf{B} , a desired alphabet size l' , and a number m with $m \geq l'$, we can define a new pseudo alphabet, \mathbf{A} , of Feldman spellings by

$$\mathbf{A} \triangleq \{\mathbf{B}_{i,m} : i = 1, 2, \dots, l'\}. \tag{2}$$

Thus, we can define a transformation as follows:

(i) pick a size ${}_0l$ for the base alphabet ${}_0\mathbf{A}$. (there is nothing else to specify since we want all the letters to have equal measure);

(ii) for each $n = 1, 2, \dots$ pick two numbers l' and m . Make ${}_n\mathbf{A}$ from ${}_{n-1}\mathbf{A}$ by letting them play the respective roles of \mathbf{A} and \mathbf{B} in (2).

Our goal is to choose, at each stage n , the numbers l' and m so as to ensure that (3.2.2) holds.

Let us first show that a cycle in a Feldman spelling is \bar{d} far away from any equal length substring of a different Feldman spelling. Fix an alphabet \mathbf{B} and, for some fixed i , let $\check{\sigma}$ denote the string Cycle_i ; so $\check{\sigma}$ is a substring of $\mathbf{B}_{i,m}$. Let $\check{\tau}$ be an equal length substring of any $\check{\tau} \triangleq \mathbf{B}_{j,m}$ where $j > i$. Then $\check{\tau}$ can consist of at most two different letters, say c and d , arranged as a long string of c followed by a long string of d .

$$\begin{array}{l} \check{\tau}: \quad cc \quad \cdots \quad cd \quad \cdots \quad d \\ \check{\sigma}: \quad \underbrace{\text{aaaa} \cdots \text{abbbb} \cdots \text{b} \cdots \text{zzzz} \cdots \text{z}}_{l \text{ substrings}} \end{array} \tag{3}$$

All the letters of \mathbf{B} appear in $\check{\sigma}$ with equal frequency. On the other hand, since $\check{\tau}$ has no more than two different letters it follows that $\bar{d}(\check{\sigma}, \check{\tau}) \geq (l-2)/l$.

Life gets only a little more complicated if we allow \mathbf{B} to be a pseudo alphabet and let d denote $\bar{d}_\varepsilon(\mathbf{B})$ for some $\varepsilon < \frac{1}{2}$. So in this context, each of a, b, c, d, \dots, z represents a pseudo letter, that is, a string of letters. What is the added complication? The string $\check{\tau}$ may or may not start at the beginning of a pseudo letter. If it does, then the situation is still as in (3) and hence $\bar{d}(\check{\sigma}, \check{\tau}) \geq [(l-2)/l]d$. On the other hand, if $\check{\tau}$ begins somewhere in the middle of a pseudo letter then the $\check{\tau}$ row of (3) is shifted horizontally by a fraction of a pseudo letter. Each pseudo letter p in $\check{\sigma}$ will see above it, in $\check{\tau}$, parts of two consecutive pseudo letters; call them q and r . At least $(l-2)/l$ -percent of the pseudo letters in $\check{\sigma}$ are equal to neither of the two pseudo letters above them; so we examine the situation when $p \neq q, r$.

Write the pseudo letter p as a concatenation of two substrings $p = \check{p}\hat{p}$ where \check{p} and \hat{p} are the parts of p which lie beneath q and r , respectively. Let \check{q} and \hat{r} denote the

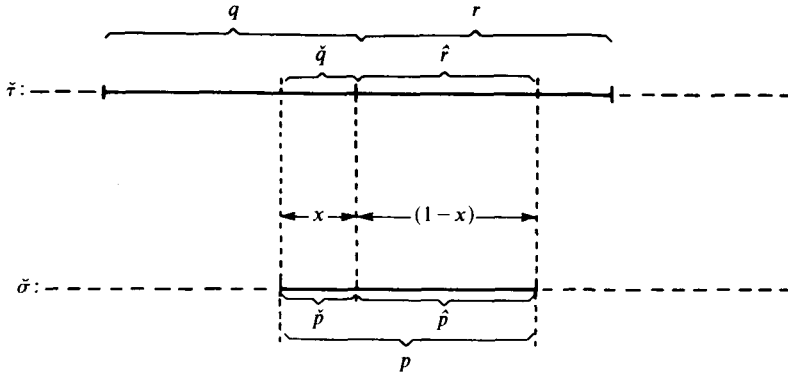


FIGURE 1

substrings of q and r lying above \check{p} and \hat{p} , respectively. Denote by x the fraction of p that q overlaps i.e. $x \triangleq \text{Len}(\check{p})/\text{Len}(p)$. Without loss of generality, $1-x > \varepsilon$. Thus $\bar{d}(\hat{p}, \hat{r}) > d$. If, in addition, $x \geq \varepsilon$ then $\bar{d}(\check{p}, \check{q}) > d$. Thus $\bar{d}(\check{p}\hat{p}, \check{q}\hat{r})$ exceeds $x \cdot d + (1-x) \cdot d = d$, unless $x < \varepsilon$. In that case, the \bar{d} distance exceeds $x \cdot 0 + (1-\varepsilon) \cdot d$. So in either instance we may conclude

$$\bar{d}(\text{Cycle}_i, \text{equal length piece of } \bar{\tau}) \geq \frac{l-2}{l} \cdot (1-\varepsilon) \cdot \bar{d}_\varepsilon(\mathbf{B}).$$

(3.2.4) FELDMAN-SPELLING THEOREM. Given ε , we can find a size $l = l(\varepsilon)$ such that for any new ε' and l' , if \mathbf{B} is a pseudo alphabet with $|\mathbf{B}| \geq l$ then for any sufficiently large m , the pseudo alphabet $\mathbf{A} \triangleq \{\mathbf{B}_{i,m} : i = 1, 2, \dots, l'\}$ satisfies

$$\bar{d}_\varepsilon(\mathbf{A}) \geq (1-\varepsilon)^2 \cdot \bar{d}_\varepsilon(\mathbf{B})$$

and, of course $|\mathbf{A}| = l'$.

Proof. Let $l(\varepsilon)$ be some integer l so that $(l-2)/l > 1-\varepsilon$. Choose a positive integer c obeying

$$\frac{c-2}{c} \cdot \frac{l-2}{l} > 1-\varepsilon. \tag{4}$$

Pick m big enough that any Feldman spelling in \mathbf{A} has at least c/ε' cycles, in the manner of (1). Take any two distinct $\bar{\sigma}, \bar{\tau} \in \mathbf{A}$. Without loss of generality, the cycles in $\bar{\sigma}$ are shorter than those of $\bar{\tau}$. Pick respective equal length substrings $\check{\sigma}$ and $\check{\tau}$ of relative length greater than ε' -percent. Then $\check{\sigma}$ is at least c cycles in length. It may, of course, start and/or end somewhere in the middle of a cycle but at least $((c-2)/c)$ -percent of its length is made up of complete cycles. Consequently,

$$\begin{aligned} \bar{d}(\check{\sigma}, \check{\tau}) &\geq \frac{c-2}{c} \cdot \bar{d}(\text{a cycle of } \check{\sigma}, \text{ equal length piece of } \bar{\tau}) \\ &\geq \frac{c-2}{c} \frac{l-2}{l} (1-\varepsilon) \cdot \bar{d}_\varepsilon(\mathbf{B}) \\ &\geq (1-\varepsilon)^2 \cdot \bar{d}_\varepsilon(\mathbf{B}) \end{aligned}$$

as desired. □

It will be convenient temporarily to denote the \mathbf{A} manufactured by the foregoing theorem as: $\mathbf{A} \triangleq \text{Feld}(\mathbf{B}, \varepsilon, \varepsilon', l')$. Note that there is a hidden parameter, m , which we can freely make as large as we desire.

Remark. For future reference, observe that since m can be made arbitrarily large, the ratio $\text{Len}(\mathbf{B})/\text{Len}(\mathbf{A})$ can be made as close to zero as desired. In particular, we can replace (4) by

$$\frac{\varepsilon' \cdot \text{Len}(\mathbf{A}) - \text{Len}(\mathbf{B})}{\varepsilon' \cdot \text{Len}(\mathbf{A})} \cdot \frac{c-2}{c} \frac{l-2}{l} > 1 - \varepsilon,$$

and replay the subsequent reasoning to conclude:

$$\frac{\varepsilon' \text{Len}(\mathbf{A}) - \text{Len}(\mathbf{B})}{\varepsilon' \text{Len}(\mathbf{A})} \cdot \bar{d}_{\varepsilon'}(\mathbf{A}) \geq (1 - \varepsilon)^2 \cdot \bar{d}_{\varepsilon'}(\mathbf{B}). \tag{5}$$

(3.2.5) *Application.* We apply the foregoing to get the transformation \mathcal{F} as follows: Pick any number $g \in (0, 1)$. Pick positive ${}_n\varepsilon \rightarrow 0$ such that $\prod_{n=0}^{\infty} (1 - {}_n\varepsilon)^2 > g$.

(i) Let ${}_n l \triangleq l({}_n\varepsilon)$, $n = 0, 1, 2, \dots$. Let ${}_0\mathbf{A}$ be an alphabet with ${}_0 l$ many letters.

(ii) Let ${}_n\mathbf{A} \triangleq \text{Feld}({}_{n-1}\mathbf{A}, {}_{n-1}\varepsilon, {}_n\varepsilon, {}_n l)$, $n = 1, 2, \dots$

Since $\bar{d}_{{}_0\varepsilon}({}_0\mathbf{A}) = 1$, we obtain the desired $\bar{d}_{{}_n\varepsilon}({}_n\mathbf{A}) > \prod_{i=0}^{n-1} (1 - {}_i\varepsilon)^2 > g$.

(3.3) *The counterexample transformation \mathcal{T} .* The transformation \mathcal{F} endowed with two additional properties will give us \mathcal{T} . Take a typical \mathcal{F} -name in the base of the n -gadget and scan along it into the future in jumps of $h(n)$ many base letters. Each jump leaps over a pseudo letter of ${}_n\mathbf{A}$. Notice that, with high frequency, the pseudo letter we are now scanning equals the one we have just finished scanning. This is because we built our $(n + 1)$ -gadget from Feldman spellings and such spellings have lots of consecutive duplications, (see (1')).

Our counterexample transformation \mathcal{T} , however, wants just the opposite: a low frequency of consecutive duplications. A frequency, in fact, going to zero as $n \rightarrow \infty$. We can ensure that the frequency at stage $n - 1$ does not exceed, say, $1/n$, by *grouping* the ${}_{n-1}\mathbf{A}$ -letters together, n at a time, and building ${}_n\mathbf{A}$ from Feldman spellings using these *groups* as ‘letters’.

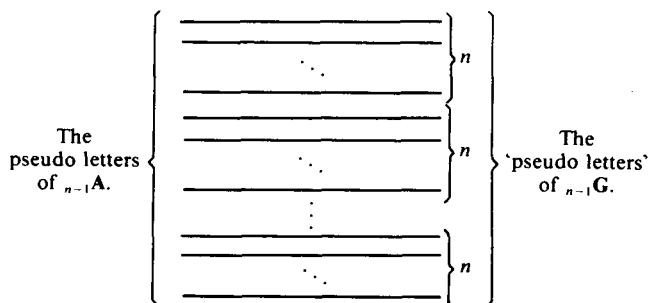


FIGURE 2. The ‘pseudo letters’ of ${}_{n-1}\mathbf{G}$ are groups with n pseudo letters of ${}_{n-1}\mathbf{A}$ to a group. We will form ${}_n\mathbf{A}$ via Feldman spellings over ${}_{n-1}\mathbf{G}$ rather than over ${}_{n-1}\mathbf{A}$.

Each Feldman spelling over the collection of $(n - 1)$ -groups will determine lots of columns in the n -gadget in the following manner. Think of different $(n - 1)$ -groups as buckets painted different colours. Each colour bucket contains n different pseudo letters from ${}_{n-1}\mathbf{A}$; each pseudo letter appears in exactly one colour of bucket. A specific Feldman spelling is a particular sequence of colours – or rather, of coloured buckets. To fabricate a n -column, we simply concatenate together the sequence of ${}_{n-1}\mathbf{A}$ -letters obtained in the following fashion: Walk along the sequence of buckets and pick, at random, an ${}_{n-1}\mathbf{A}$ -letter from each. Evidently the probability of picking the exact same ${}_{n-1}\mathbf{A}$ -letter from two adjacent buckets is zero, if their colours differ, and $1/n$, if they happen to have the same colour; hence this probability is no more than $1/n$.

This ‘grouping’ idea will necessitate a redefinition of the meaning of ‘pseudo letter’. We proceed to the details.

(3.3.1) *Definition.* A pseudo alphabet \mathbf{B} is still a collection of pseudo letters but now a *pseudo letter* $\xi \in \mathbf{B}$ is to be a set of equal length words over the base alphabet. Each word $\bar{\tau} \in \xi$ is a *member* of ξ and we ask that each pseudo letter of $\mathbf{B} = \{\xi_1, \dots, \xi_r\}$ have the same number of members. We extend the definition of $\bar{d}_\epsilon(\cdot, \cdot)$ to mean:

$$\bar{d}_\epsilon(\xi, \xi') \triangleq \inf \{ \bar{d}_\epsilon(\bar{\sigma}, \bar{\tau}) : \bar{\sigma} \in \xi, \bar{\tau} \in \xi' \},$$

whereas $\bar{d}_\epsilon(\cdot)$ simply retains its meaning of $\bar{d}_\epsilon(\mathbf{B}) \triangleq \inf \{ \bar{d}_\epsilon(\xi, \xi') : \xi, \xi' \in \mathbf{B}, \xi \neq \xi' \}$.

We would like to still be able to take Feldman spellings over such a \mathbf{B} . Formula (1) defined the pseudo letter $\mathbf{B}_{i,m}$ as a long concatenation of the pseudo letters of \mathbf{B} . But now pseudo letters are sets of words – so $\mathbf{B}_{i,m}$ can still be defined by (1) once we say what we mean by the concatenation of two *sets* of words. We concatenate independently: the concatenation of the two pseudo letters $\{\bar{\sigma}_1, \dots, \bar{\sigma}_I\}$ with $\{\bar{\tau}_1, \dots, \bar{\tau}_J\}$ we define to be the pseudo letter

$$\{ \bar{\sigma}_i \bar{\tau}_j : i \in [1, I] \text{ and } j \in [1, J] \}.$$

One may check that theorem (3.2.4) continues to hold.

We can now elaborate on the grouping in figure 2. Each horizontal line represents a pseudo letter of ${}_{n-1}\mathbf{A}$, that is, a set of words. Writing ${}_{n-1}\mathbf{A}$ as $\{\xi_1, \xi_2, \dots, \xi_r\}$, we mean by grouping n at a time that the first pseudo letter of ${}_{n-1}\mathbf{G}$ be $\bigcup_{i=1}^n \xi_i$; that the second be $\bigcup_{i=n+1}^{2n} \xi_i$ and so on.

(3.3.2) *Application.* We construct a transformation \mathcal{F} and pseudo alphabets $\{ {}_n\mathbf{A} \}_n$ with the old property $\forall \epsilon, \forall_{\text{large } n} : \bar{d}_\epsilon({}_n\mathbf{A}) > g$ as well as the ‘infrequency of consecutive duplications’ property. This will be stated precisely after the construction of \mathcal{F} , in (3.3.5c(i)). As before, we fix a $g \in (0, 1)$ and ${}_\epsilon \rightarrow 0$ such that $\prod_{n=0}^\infty (1 - \epsilon)^2 > g$.

(i) Set ${}_n l \triangleq (n + 1) \cdot l(\epsilon)$, $n = 0, 1, 2, \dots$. Pick some base alphabet ${}_0\mathbf{A} = \{a, b, \dots, z\}$ with ${}_0 l$ many letters. For the purposes of this induction, regard ${}_0\mathbf{A}$ also as the pseudo alphabet $\{\{a\}, \{b\}, \dots, \{z\}\}$.

(ii) Take the pseudo letters of ${}_{n-1}\mathbf{A}$ and group them n at a time to form the pseudo letters of ${}_{n-1}\mathbf{G}$, as in figure 2. Thus $|{}_{n-1}\mathbf{G}| = l({}_{n-1}\epsilon)$ and, of course,

$\bar{d}_\varepsilon({}_{n-1}\mathbf{G}) \geq \bar{d}_\varepsilon({}_{n-1}\mathbf{A})$ for any ε . Let

$${}_n\mathbf{I} \triangleq \text{Feld}({}_{n-1}\mathbf{G}, {}_{n-1}\varepsilon, {}_n\varepsilon, {}_n l)$$

as n ranges over $1, 2, \dots$

(iii) Finally, let ${}_n\mathbf{A} \triangleq {}_n\mathbf{I}$.

Notice that the transformation just constructed, $\check{\mathcal{T}}$, is certainly not a \mathcal{K} -automorphism; its powers are not even ergodic, since each level of the n -gadget is invariant under $\check{\mathcal{T}}^{h(n)}$. We now modify the above cutting and stacking procedure in what has become a standard way to force \mathcal{K} -ness of a transformation. This method was introduced by Ornstein and Shields in their classic paper [6].

(3.3.3) *Springs.* The idea is to put a small amount of randomness into the location of the $(n-1)$ -blocks within the n -blocks. We achieve this by adding in, at each stage n , a little filler material at the ends of Feldman spellings in a random way. This will involve retaining steps (i) and (ii) from above but replacing (iii) by (iii') below.

Pick two new letters S and C not appearing in ${}_0\mathbf{A}$. They will form the filler material that we will concatenate independently to each end of the pseudo letters of ${}_n\mathbf{I}$ to form the pseudo letters of ${}_n\mathbf{A}$. Denote the pseudo letters of ${}_n\mathbf{I}$ as ${}_n\mathbf{I} = \{\xi_1, \dots, \xi_j, \dots, \xi_J\}$. Now replace (iii) by (iii'):

(iii') Let ${}_n\mathbf{A} \triangleq \{\psi_1, \dots, \psi_j, \dots, \psi_J\}$ where

$$\psi_j \triangleq \underbrace{\{\text{SS} \dots \text{S}\}}_i \bar{\sigma} \underbrace{\{\text{CC} \dots \text{C}\}}_k : \text{where } \bar{\sigma} \in \xi_j, i \in [0, h(n-1)) \text{ and } i+k = h(n-1)\}.$$

The S's are the start of the spring and the C's are the spring complement. The number of S's plus the number of C's we call the length of the n -spring. We only need that this length should go to infinity as $n \rightarrow \infty$. However, we chose $\text{Len}(n\text{-spring}) = h(n-1)$ for convenience so as to have the ratio $h(n)/h(n-1)$ always an integer.

To ensure that our cutting and stacking procedure (i)(ii)(iii') does indeed produce a transformation \mathcal{T} which has not lost the properties of $\check{\mathcal{T}}$, we need to check two items: First, we need to show that the total amount of filler we have added is finite. As noted in the remark following the Feldman-spelling theorem, the pseudo alphabet created by that theorem can be made arbitrarily long. Hence, for each n we can have made the ratio $h(n-1)/h(n) \stackrel{\text{note}}{=} \mu(n\text{-spring})/\mu(n\text{-gadget})$ as small as desired, hence summable. Consequently

$$\sum_{n=1}^{\infty} \frac{\mu(\text{material added at stage } (n-1))}{\mu(n\text{-gadget})}$$

is finite, and hence we have indeed defined a transformation \mathcal{T} on a probability space.

Secondly, we also need check that we have preserved the \bar{d}_ε property. This too follows from (5) because, essentially,

$$\begin{aligned} \bar{d}_{\varepsilon'}({}_n\mathbf{A}) &\geq \frac{\varepsilon' h(n) - \text{Len}(n\text{-spring})}{\varepsilon' h(n)} \cdot \bar{d}_{\varepsilon'}({}_n\mathbf{I}) = \frac{\varepsilon' h(n) - h(n-1)}{\varepsilon' h(n)} \cdot \bar{d}_{\varepsilon'}({}_n\mathbf{I}) \\ &\geq (1 - \varepsilon)^2 \bar{d}_\varepsilon({}_{n-1}\mathbf{A}) \end{aligned}$$

where ε' and ε represent ${}_n\varepsilon$ and ${}_{n-1}\varepsilon$, respectively.

(3.3.4) *How general is the construction?* Our construction is flexible enough to build \mathcal{T} in the following general way: We first choose a number $\text{Gap} \in (0, \frac{1}{3})$ where $3 \cdot \text{Gap}$ will play the role of g . Then freeze, once and for all, the numbers $\{n\varepsilon\}$ and then the numbers $\{nI\}$ as we chose them in (3.3.2). The $\{n\varepsilon\}$ and $\{nI\}$, as we saw in the Feldman-spelling theorem, determine an *a priori lower bound*, but no upper bound, on the value of the hidden parameter used at stage n – which is to say, on the ratio $h(n+1)/h(n)$. Thus, we could ask that these ratios also dominate any additional sequence of lower bounds obtained in some fashion or other. In particular, given a function $M: \mathbb{N} \rightarrow \mathbb{N}$ we can, at each stage n , make the $(n+1)$ -gadget tall enough, that is, make $h(n+1)/h(n)$ large enough, so that most points in the $(n+1)$ -gadget are in some n -letter and are at least $M(h(n))$ many n -letters away from the top of the $(n+1)$ -gadget. We will see a formal statement of this in (3.3.5c(ii)) below.

It will not be until § 7 that we define this function $M(\cdot)$ and so we must be careful not to use implicitly any property of \mathcal{T} – such as how soon the ergodic theorem kicks in on a particular set – in the definition of $M(\cdot)$.

Having laboriously constructed \mathcal{T} , we can now efface from our memories the details of its construction and just retain the following properties:

(3.3.5) *Relevant properties.* Given any ‘lower bound’ function $M: \mathbb{N} \rightarrow \mathbb{N}$ we can build a Kolmogorov automorphism \mathcal{T} (\mathcal{H} -ness to be verified in the next section) and a sequence of numbers $\{nM\}$, where $nM \triangleq M(h(n))$, satisfying the collection of properties below.

At stage n the columns of the n -gadget all have equal measure. We collect them into equal sized subsets to form the pseudo alphabet ${}_n\mathbf{A}$. Agree henceforth to refer to ${}_n\mathbf{A}$ as the n -alphabet and to its pseudo letters as n -letters or just letters. For any $\tau \in n$ -gadget, we let ${}_n\mathbf{A}\tau$ denote the n -letter that τ is in; we let $({}_n\mathbf{A}\tau)|_k$ mean the k th n -letter one sees when scanning into the future on τ . Now \mathcal{T} satisfies:

- (a) $h(n) \rightarrow \infty$, as $n \rightarrow \infty$.
- (b) $\forall \varepsilon, \forall \text{large } n: \bar{d}_\varepsilon(\xi, \xi') > 3 \cdot \text{Gap}$, for any two distinct n -letters $\xi, \xi' \in {}_n\mathbf{A}$.
- (c) *Infrequency of duplications.* $\forall \delta, \forall \text{large } n$ there is a set $U \subset X$ with $\mu(U^c) < \delta$ such that for each $\tau \in U$:

(i) $({}_n\mathbf{A}\tau)|_0^M$ has less than δ -percent consecutive duplications. This means that for at least $(1 - \delta)$ -percent of $k \in [0, M]$: $({}_n\mathbf{A}\tau)|_k \neq ({}_n\mathbf{A}\tau)|_{k+1}$.

(ii) Scanning τ from time-zero forward, one sees (at least) nM many contiguous n -letters. In other words, there is an $i' \in [0, h(n)]$ for which $\mathcal{T}^{i'+kh(n)}(\tau)$ is a point in the base of the n -gadget for every $k \in [0, nM]$.

4. \mathcal{T} is a Kolmogorov automorphism

$\mathbf{A} \triangleq {}_0\mathbf{A} \cup \{\mathbf{S}, \mathbf{C}\}$ is a generating partition for \mathcal{T} . The exposition of our proof of \mathcal{H} -ness is as in [8] with a slight modification to handle the grouping. We verify the usual property:

For all word lengths s and any $\varepsilon, \forall \text{large } g$, and for any m :

$$\bigvee_{i=-g-m}^{-g-1} \mathcal{T}^i \mathbf{A} \perp^{4\varepsilon} \bigvee_{i=0}^{s-1} \mathcal{T}^i \mathbf{A}. \tag{6}$$

This says that the distribution of s -words is essentially independent, 4ϵ -independent, from the remote past – that is, when separated from the past by a sufficiently large gap g . By ‘the distribution of s -words’ we mean the tuple of numbers $\mu\{\tau: \tau|_0^s = \check{\sigma}\}$ as $\check{\sigma}$ ranges over all s -words in some fixed order. The distribution of s -words with τ restricted to some set SET will be indicated by $\text{distr}(\tau \in SET)$.

Proof. Choose n sufficiently large that $\mu(X') > 1 - \epsilon^2$ where the set X' is the $(n-1)$ -stack minus its last s levels. It will suffice to establish

$$\bigvee_{-g-m}^{-g-1} \mathcal{F}^i \mathbf{A}/X' \perp^{2\epsilon} \bigvee_0^{s-1} \mathcal{F}^i \mathbf{A}/X', \tag{6'}$$

and so we restrict our attention to those names τ in X' .

Note that $\text{distr}(\tau \in X') = \text{distr}(\tau \in X' \ \& \ {}_n \mathbf{A}\tau = \xi)$ for any particular n -letter ξ . This follows since the Feldman spelling of ξ sees each $(n-1)$ -group with equal probability and, consequently, each column of the $(n-1)$ -gadget with equal probability.

Pick N so that $\text{Len}(N\text{-spring}) \triangleq h(N-1) \gg h(n)$, then fix any $g > h(N)$. Define the measurable functions $N(\tau)$, $l(\tau)$, $n(\tau)$ as follows: $N(\tau)$ is the first coordinate of the N -block containing τ_0 ; $l(\tau)$ is the length of this n -block’s forward spring; $n(\tau)$ is the distance between τ_0 and the beginning of the n -block containing time zero.

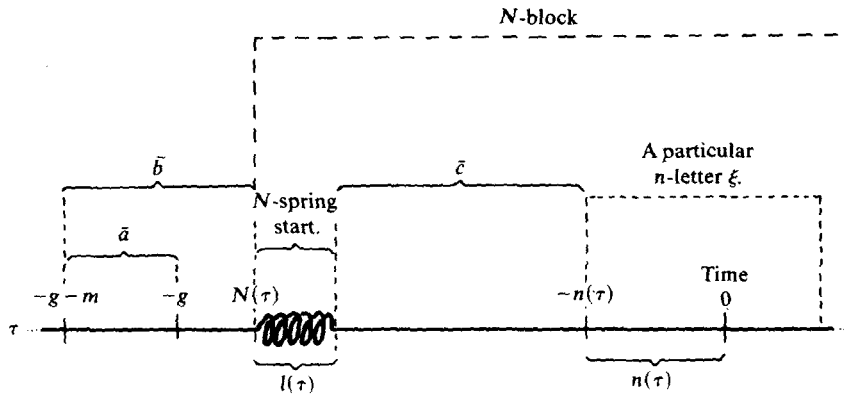


FIGURE 3. The quantities shown below the name, $n(\tau)$ and $l(\tau)$, are lengths. Indices 0, $-n(\tau)$, $N(\tau)$, $-g$ and $-g-m$, are written above the name τ . Finally, \bar{a} , \bar{b} , and \bar{c} are substrings.

For each m -word \bar{a} , let the symbol \bar{a} also denote the set $\{\tau \in X': \tau|_{-g-m}^{-g} = \bar{a}\}$. We can split \bar{a} up into subsets $\{\bar{a}_k\}$ so that $\tau \in \bar{a}_k$ means:

- $N(\tau)$ is some fixed number $N_0 \in \mathbb{Z}_{-g}^0$;
- $\tau|_{-g-m}^{N_0}$ is some fixed string \bar{b} ;
- $\tau|_{N_0+l(\tau)}^{-n(\tau)}$ is some fixed string \bar{c} ; and
- ${}_n \mathbf{A}\tau$ is some fixed n -letter ξ .

As τ ranges over an \bar{a}_k , the function $l(\tau)$ varies uniformly and consequently $n(\tau)$ varies uniformly over that subset of $\mathbb{Z}_0^{h(n)-s}$ consistent with the constraints

$$l(\tau) + \text{Len}(\bar{c}) + n(\tau) = |N_0|$$

and, of course, $l(\tau) \in \mathbb{Z}_0^{h(N-1)}$. Agree to call an \bar{a}_k good if the constraints are vacuous

and $n(\tau)$ varies uniformly on $|_0^{h(n)-s}$. As τ ranges over this set \bar{a}_k , we always see ξ at time zero. But which member of ξ we see is independent from everything in sight; in particular from $l(\tau)$ and hence from $n(\tau)$. So, for a good \bar{a}_k

$$\text{distr}(\tau \in \bar{a}_k) = \text{distr}(\tau \in X' \ \& \ nA\tau = \xi) = \text{distr}(\tau \in X').$$

If all \bar{a}_k were good then we could conclude, by averaging over the $\{\bar{a}_k\}$, that $\text{distr}(\tau \in \bar{a}) = \text{distr}(\tau \in X')$. This would be (6') with actual independence.

The union of all bad \bar{a}_k is contained in the complement, with respect to X' , of the set

$$\{\tau \in X' : h(n) < l(\tau) + n(\tau) < h(N - 1)\}.$$

We can have made N sufficiently large that this complement has measure less than ε^2 . Hence (6') follows. □

5. The code

We now assume, for the sake of contradiction, that we have an isomorphism $\varphi : \Lambda \rightarrow X$ from some simple independent skew $\mathcal{S} = \mathcal{B} \otimes (\mathcal{X}/\mathbf{H}, \text{Id}/\mathbf{T})$ to our \mathcal{T} i.e. $\mathcal{T}\varphi = \varphi\mathcal{S}$ where, we recall, $\Lambda \triangleq \Omega_{\mathcal{B}} \times \Omega_{\mathcal{X}}$ is the space on which \mathcal{S} acts. Recall that the product partition $\mathbf{P}' \triangleq (\mathbf{H}, \mathbf{T}) \times \mathbf{a}$ a generating partition for \mathcal{X}' lives on the product space Λ and generates under \mathcal{S} . Since \mathcal{S} and \mathcal{T} are isomorphic, we can think of \mathcal{T} 's generating partition \mathbf{A} as also living on Λ . Hence, given any small number δ , we can find a k such that $\bigvee_{i=-k}^k \mathcal{S}^i \mathbf{P}' \supset_{\delta} \mathbf{A}$. This allows us to define a finite code \mathcal{C} which well approximates φ . More precisely, we can define a map $\mathcal{C} : \{\mathbf{P}'\text{-}|_{-k}^{k+1}\text{-words}\} \rightarrow \mathbf{A}$ and a set $BAD \subset \Lambda$ of measure less than δ , such that $\lambda \in \Lambda \sim BAD$ implies $\mathcal{C}(\lambda|_{-k}^{k+1}) = (\varphi\lambda)|_0$. By the ergodic theorem, the orbit of any point λ hits BAD with frequency $\mu(BAD) < \delta$ and so our code \mathcal{C} makes \bar{d} errors less than δ -percent of the time. By this, one means that δ exceeds the frequency of those $i \in \mathbb{N}$ for which $\mathcal{C}(\lambda|_{i-k}^{i+k+1}) \neq (\varphi\lambda)|_i$.

(5.1) *A convention.* To make a 'guess' at the letter $(\varphi\lambda)|_i$, our code needs to look at $\lambda|_{i-k}^{i+k+1}$. If we permit our code to see only some substring $\lambda|_0^m$ of the name λ , then our code can guess at $(\varphi\lambda)|_k^{m-k}$. So we could certainly view our code as a mapping from $\mathcal{S}\text{-}|_0^m\text{-words}$ to $\mathcal{T}\text{-}|_k^{m-k}\text{-words}$, that is, from the set $\mathcal{S}|_0^m$ to $\mathcal{T}|_k^{m-k}$. In fact, although our code, when applied to $\lambda|_0^m$, cannot guess at the first or last k letters of $(\varphi\lambda)|_0^m$, we can think of \mathcal{C} as a mapping $\mathcal{S}|_0^m \rightarrow \mathcal{T}|_0^m$ by just assigning the first and last k letters in any fixed way. Moreover, if $m \gg k$ - so much so that $(2k/m) + \mu(BAD) < \delta$, then we can view our code as mapping $\mathcal{S}\text{-}m\text{-words}$ to $\mathcal{T}\text{-}m\text{-words}$ and still having a \bar{d} error less than δ . This is a standard convention.

In the above discussion, \mathcal{S} could have been just any transformation. Since, in our case, \mathcal{S} is a skew product $\mathcal{B} \otimes (\mathcal{X}/\mathbf{H}, \text{Id}/\mathbf{T})$ it will be convenient to view \mathcal{C} , rather, as a mapping

$$\mathcal{C} : \mathcal{B}|_0^m \times \mathcal{X}|_0^m \rightarrow \mathcal{T}|_0^m.$$

We fix now for the duration of this paper a code \mathcal{C} whose \bar{d} error is less than the number Gap from the construction of \mathcal{T} . The above convention is in force. Hence, whenever we speak of applying the code to some finite string, it is tacitly to be

assumed that this string's length is much bigger than the code length k . Usually the role of m above will be played by $h(n)$ or $\varepsilon \cdot h(n)$ where n is chosen 'sufficiently large'.

(5.2) *Definition.* Agree to say that the substrings $\lambda|_a^b$ codes well if $\bar{d}(\mathcal{C}(\lambda|_a^b), (\varphi\lambda)|_a^b) < \text{Gap}$. For any positive ε we say that $\lambda|_a^b$ codes well on any ε -percent if, in addition, each substring $\lambda|_i^{i+\varepsilon(b-a)}$ codes well for each $i: a \leq i < i + \varepsilon(b-a) \leq b$.

(5.3) STANDARD CODING LEMMA. $\forall \varepsilon, \delta > 0, \forall_{\text{large } m}$, there is a set $U \subset \Lambda$ with $\mu(U^c) < \delta$ such that $\lambda|_0^m$ codes well on any ε -percent, for each $\lambda \in U$.

Proof. Standard. □

(5.4) *Definition.* Given any point $\tau \in X$ we can scan forward from time-zero along the name and find indices $0 \leq i_0 < i_1 < \dots$ so that i_k is the smallest value $i > i_{k-1}$ for which $\mathcal{T}^i \tau \in \text{base}$ of the n -gadget. Thus $\tau|_{i_k}^{i_k+h(n)}$ is a single column in the n -gadget and hence $i_k + h(n) \leq i_{k+1}$. Let $({}_n\mathbf{A}\tau)|_k$ denote the n -letter of ${}_n\mathbf{A}$ of which $\tau|_{i_k}^{i_k+h(n)}$ is a member. Thus $({}_n\mathbf{A}\tau)|_0^\infty$ denotes a half-infinite sequence of n -letters; we call this the n -spelling of τ .

If $\tau \in n$ -gadget, let ${}_n\mathbf{A}\tau$ denote the n -letter that τ is in. To make sure we have the notation straight, note that ${}_n\mathbf{A}\tau = ({}_n\mathbf{A}\tau)|_0$ if i_0 happens to equal 0; otherwise ${}_n\mathbf{A}\tau = ({}_n\mathbf{A}\tau)|_{-1}$.

Now we can use the isomorphism φ to lift all these concepts to the space Λ . Thus $\langle \zeta_\beta \rangle \in n$ -gadget means $\varphi(\zeta_\beta) \in n$ -gadget; $({}_n\mathbf{A}\langle \zeta_\beta \rangle)|_0^\infty$ means $({}_n\mathbf{A}\varphi(\zeta_\beta))|_0^\infty$, etc.

6. Oracles

Our goal is to show, granted n large, that for most points $\langle \zeta_\beta \rangle$ the n -letter at time zero, ${}_n\mathbf{A}\langle \zeta_\beta \rangle$, depends only on ζ and not on β .

(6.1) *Definition.* For any fixed, implicit n we say that $\zeta \in \Omega_{\mathcal{X}}$ is *oracular* or is an *oracle* if there exists a letter $\xi \in {}_n\mathbf{A}$ such that the set $\{\beta \in \Omega_{\mathcal{B}}: {}_n\mathbf{A}\langle \zeta_\beta \rangle = \xi\}$ has measure exceeding $\frac{1}{2}$.

We can now define a map $\mathring{\mathbf{A}}$, an *oracle function* from $\Omega_{\mathcal{X}}$ to ${}_n\mathbf{A} \cup \{\text{'undefined'}\}$, by letting $\mathring{\mathbf{A}}\zeta \triangleq \xi$ whenever ξ as above exists.

Finally, we call a pair $\langle \zeta_\beta \rangle$ *oracular* or an *oracle-pair* if ${}_n\mathbf{A}\langle \zeta_\beta \rangle = \mathring{\mathbf{A}}\zeta$.

(6.2) ORACLE-PAIR ABUNDANCE THEOREM. $\forall \delta, \forall_{\text{large } n}$ we can find a large oracular set $U \subset \Lambda$ with $\mu(U) > 1 - \delta$, such that any $\langle \zeta_\beta \rangle \in U$ is an oracle-pair.

Proof. Fix any $\varepsilon < (\delta/100)^{100}$. It suffices to show that for all large n there exists a good set $G \subset \Omega_{\mathcal{X}}$ with $\mu(G) > 1 - (\delta/2)$, such that each $\zeta \in G$ has a corresponding set $B_\zeta \subset \Omega_{\mathcal{B}}$ with $\mu(B_\zeta) > 1 - (\delta/2)$, such that $\langle \zeta_\beta \rangle$ is oracular whenever $\beta \in B_\zeta$. Since, without loss of generality, $1 - (\delta/2)$ exceeds the $\frac{1}{2}$ in the definition of oracle, it suffices to show that ${}_n\mathbf{A}\langle \zeta_\beta \rangle = {}_n\mathbf{A}\langle \zeta_{\beta'} \rangle$ for all $\beta, \beta' \in B_\zeta$.

If we pick n large, our standard coding lemma will give us a set $V \subset \Lambda$ with $\mu(V^c) < \varepsilon$, such that when $\lambda \in V$, both $\lambda|_{-\varepsilon h(n)}^0$ and $\lambda|_0^{\varepsilon h(n)}$ code well. We may ask that ${}_n\varepsilon < \varepsilon$. Now only a *tiny* percentage of $\lambda \in \Lambda$ are such that $\varphi\lambda$ is not in the n -gadget. Of the remaining λ , exactly 2ε -percent have $(\varphi\lambda)|_{-\varepsilon h(n)}^{\varepsilon h(n)}$ crossing an n -block

boundary. We discard both kinds of bad λ from V and consequently have

$$\mu(V^c) < \varepsilon + \text{tiny} + 2\varepsilon \stackrel{\text{WLOG}}{<} 4\varepsilon \stackrel{\text{def}}{=} \delta'.$$

We have now guaranteed, whenever $\lambda \in V$, that $(\varphi\lambda)|_{-\varepsilon h(n)}^{\varepsilon h(n)}$ is a substring of some member of some n -letter.

For each $\zeta \in \Omega_{\mathcal{X}}$ let B'_ζ denote $\{\beta: \langle \zeta \rangle \in V\}$. Via a Chebyshev argument, at most $\sqrt{\delta'}$ -percent of names ζ are so unpopular that $\mu(B'_\zeta)$ is less than $1 - \sqrt{\delta'}$. We let G be the set of popular ζ and have that

$$\mu(G) > 1 - \sqrt{\delta'} > 1 - \frac{\delta}{2}$$

by our upper bound on ε .

Fix some $\zeta \in G$ for the remainder of this proof. Suppose $\beta, \beta' \in B'_\zeta$ have the same prefix i.e. $\beta|_{-\varepsilon h(n)}^0 = \beta'|_{-\varepsilon h(n)}^0$. Then ${}_n\mathbf{A}\langle \zeta \rangle = {}_n\mathbf{A}\langle \beta \rangle$ because $\langle \zeta \rangle|_{-\varepsilon h(n)}^0$ and $\langle \beta \rangle|_{-\varepsilon h(n)}^0$ are equal and they each code well i.e. code to $(\varphi\langle \beta \rangle)|_{-\varepsilon h(n)}^0$ and $(\varphi\langle \beta' \rangle)|_{-\varepsilon h(n)}^0$, respectively, to within a \bar{d} error of Gap. Since $\varepsilon > {}_n\varepsilon$, we know that distinct n -letters are at least 3 Gap apart in \bar{d}_ε ; hence

$${}_n\mathbf{A}\langle \beta \rangle = {}_n\mathbf{A}\langle \beta' \rangle. \tag{7}$$

Or, this would still hold if, instead, β and β' had the same suffix i.e. $\beta|_0^{\varepsilon h(n)} = \beta'|_0^{\varepsilon h(n)}$. Consequently, if $\beta, \beta', \beta'' \in B'_\zeta$ are such that β and β' have the same prefix, and β' and β'' have the same suffix, then ${}_n\mathbf{A}\langle \beta \rangle = {}_n\mathbf{A}\langle \beta'' \rangle$.

The independence of the \mathcal{B} -process will now complete the proof. For each prefix $\check{\rho} \in \mathcal{B}|_{-\varepsilon h(n)}^0$, define a corresponding ‘suffix set’

$$\check{F}_\rho \triangleq \{\beta|_0^{\varepsilon h(n)}: \beta \in B'_\zeta \text{ and } \beta|_{-\varepsilon h(n)}^0 = \check{\rho}\}.$$

We can view \check{F}_ρ both as a union of cylinder sets and as a collection of strings. There must exist at least one prefix such that the measure of its suffix set is at least as large as the average value of $\mu(\check{F}_\rho)$ as $\check{\rho}$ ranges over all possible prefixes. Thus, we can fix a $\check{\rho}$ such that $\mu(\check{F}_\rho) \geq \mu(B'_\zeta)$. Consequently, the measure of the set

$$B_\zeta \triangleq B'_\zeta \cap \{\beta: \beta|_0^{\varepsilon h(n)} \in \check{F}_\rho\}$$

exceeds $1 - (\delta/2)$ by our upper bound on ε . Moreover, by the idea contained in the remark following (7), we know that the function $\beta \mapsto {}_n\mathbf{A}\langle \beta \rangle$ is constant on B_ζ . □

7. Contradiction from random overlap

From § 3 on assembly we know that:

For any small δ , $\forall_{\text{large } n}$ and for any number m , we can find a set $U \subset X$, with $\mu(U^c) < \delta$, such that any point $\tau \in U$ satisfies: the frequency of consecutive duplicate n -letters in $({}_n\mathbf{A}\tau)|_0^m$ is less than δ .

Via our isomorphism φ we can view this as a statement about the space Λ ; just replace X by Λ and τ by $\langle \zeta \rangle$. By using a Chebyshev argument we can write the statement thus:

LEMMA. *For all ‘most’, $\forall_{\text{large } n}$, for all m : Most ζ are such that for most β : Most n -letters in $({}_n\mathbf{A}\langle \beta \rangle)|_0^m$ do not equal their successor.*

In the overview of § 2 we proposed to achieve the negation of this by examining how two randomly chosen Bernoulli names β and ρ would read a common \mathcal{X} -name; how would the two parsings (see following definition) of ζ induced by $({}_n\mathbf{A}(\zeta_\beta))|_0^\infty$ and $({}_n\mathbf{A}(\zeta_\rho))|_0^\infty$ interact? To be more precise:

Definition. Given any point $\lambda = \langle \zeta \rangle$ we can scan forward from time-zero along the name and find indices $0 \leq i_0 < i_1 < \dots$ so that i_k is the smallest value $i > i_{k-1}$ for which $\mathcal{S}^i \langle \zeta \rangle \in \text{base of } n\text{-gadget}$. Letting $\tau \triangleq \varphi(\lambda)$, we see that $\tau|_{i_k}^{i_k+h(n)}$ is a single column in the n -gadget and that $i_k + h(n) \leq i_{k+1}$. In fact, the (possible empty) word $\tau|_{i_k+h(n)}^{i_{k+1}}$ consists only of springs added at stage n and greater. Define the ‘half open’ interval of integers

$$I_k = I_k(\lambda) \triangleq [\text{Weight}(\beta|_0^{i_k}), \text{Weight}(\beta|_0^{i_k+h(n)})].$$

The purpose of I_k is this: our finite code makes a guess at $\tau|_{i_k}^{i_k+h(n)}$ by looking at $\beta|_{[i_k, i_k+h(n))}$ and at $\zeta|_{I_k}$. Thus, if some index $w \in I_k$ happens to be such that $\mathcal{X}^w \zeta$ is an oracle, then this oracle makes a prediction about $({}_n\mathbf{A}(\zeta_\beta))|_k$. So we think of the $\{I_k\}_{k=0}^\infty$ as *parsing* the name ζ into pieces; the k 'th piece associated with the k 'th n -letter of $\langle \zeta \rangle$.

Our discussions will consider two points with the same \mathcal{X} -component, $\langle \zeta \rangle$ and $\langle \xi \rangle$. We shall use I_k and J_l to denote, respectively, $I_k(\langle \zeta \rangle)$ and $J_l(\langle \xi \rangle)$.

When n is large and, in consequence, springs are rare, it is often the case that the n -letters along λ near time-zero occur in a long contiguous piece i.e. that $i_k = i_0 + kh(n)$ for $k = 1, 2, \dots$ upto some large value. Thus, although the *isomorphism* φ determines $i_0 = i_0(\lambda)$, it is β that determines everything else about the $\{I_k\}_{k=1}^\infty$. It is a useful viewpoint, then, to think of β as directing the parsing of ζ . Consequently, if we randomly pick two Bernoulli names β and ρ , we might expect that their mutual ignorance would force their respective parsings $\{I_k\}$ and $\{J_l\}$ of ζ , to be oft out of phase; that is, for a high density of indices $k \in \mathbb{N}$, that the intervals I_k and I_{k+1} each overlap some interval J_l in a reasonably large percentage. If oracles are frequent,

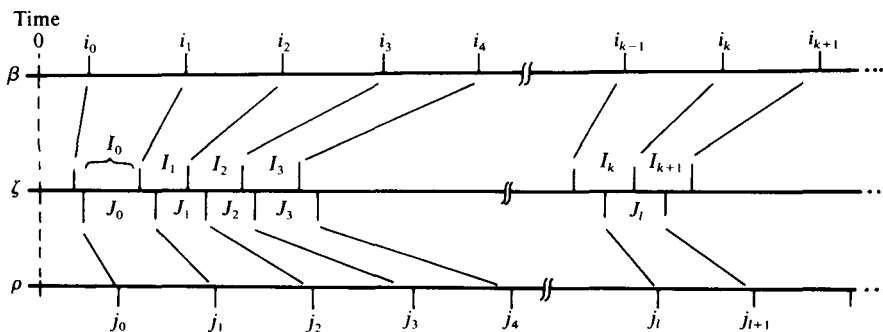


FIGURE 4. The two Bernoulli names β and ρ , as well as the \mathcal{X} -name ζ , are shown starting at time zero. The ‘hashmarks’ $\{i_k\}_{k=0}^\infty$ marked on β determine a parsing of ζ into pieces $\{I_k\}_{k=0}^\infty$ whose length is, on average, $h(n)\mu(\mathbf{H})$. Similarly the hashmarks $\{j_l\}_{l=0}^\infty$ produce a parsing $\{J_l\}_{l=0}^\infty$ of ξ . Eventually these parsings get somewhat out of phase, as shown on the right.

they will often fall in such overlaps. This makes it probable that

$$({}_n\mathbf{A}(\xi_\beta))|_k = ({}_n\mathbf{A}(\xi_\rho))|_i = ({}_n\mathbf{A}(\xi_\beta))|_{k+1},$$

which would contradict the lemma stated informally at the beginning of this section. But the overview glossed over one big headache: *springs*. After all, β and ρ might start reading in phase. And it takes a *long* time to get out of phase, since most \mathcal{B} - $h(n)$ -words have weight within a few percent of $h(n) \cdot \mu(\mathbf{H})$. Inevitably, though, they do get out of phase – unless some pesky spring comes along and syncs them up again. For remember, springs are dropped down by the isomorphism φ . And although β and ρ , being independent Bernoulli names, must each be uninformed about the other, there is nothing to prevent φ from being maliciously intelligent – it does not have to satisfy randomness requirements.

Since we cannot outsmart it, a plausible remedy is to handicap φ by giving it too little spring, at stage n , to work with. But where we determine spring infrequency – the numbers $\{\mu(X \sim (n\text{-gadget}))\}_n$ as a function of n , is at assemblytime. It is at that time we need to know how long (how many n -letters) it typically takes the n -spellings of two points $\langle \xi_\beta \rangle$ and $\langle \xi_\rho \rangle$ to get out of phase. Unfortunately, there are two pieces of information, unknown to us at assemblytime, which may influence this number: the function $\lambda \mapsto i_0(\lambda)$ determined by the isomorphism φ ; and the measure of \mathbf{H} , the skewing set. Somehow, we need to decouple these two factors from how long it takes to get out of phase.

(7.1) *Oracles revisited.* We show that oracles which predict correctly simultaneously for both $\langle \xi_\beta \rangle$ and $\langle \xi_\rho \rangle$, occur frequently along ζ . For any implicit n , any m , and any $\lambda = \langle \xi_\beta \rangle \in \Lambda$, let

$$[\text{Freq. of oracles on } |_0^m](\lambda) \triangleq \frac{1}{m} |\{i \in |_0^m : \mathcal{S}^i \lambda \text{ is an oracle-pair}\}|$$

$$[\text{Freq. of seeing } \mathbf{H} \text{ on } |_0^m](\lambda) \triangleq \frac{1}{m} |\{i \in |_0^m : \mathcal{S}^i \lambda \in \mathbf{H} \times \Omega_{\mathcal{X}}\}|.$$

Stated differently, the righthand side of the second definition is

$$\frac{1}{m} |\{i \in |_0^m : \beta_i = \mathbf{H}\}|.$$

(7.1.1) LEMMA. $\forall \delta', \forall_{\text{large}} n, \forall m$ the sets $U', V \subset \Lambda$ each have measure greater than $1 - \delta'$ where

$$U' \triangleq \{\lambda : [\text{Freq. of oracles on } |_0^{m \cdot h(n)}](\lambda) > 1 - \delta'\}; \tag{8}$$

$$V \triangleq \{\lambda : [\text{Freq. of seeing } \mathbf{H} \text{ on } |_0^{m \cdot h(n)}](\lambda) > \mu(\mathbf{H})/2\}. \tag{9}$$

Proof. Given δ' , for each large n the oracle-pair abundance theorem provides us with an oracular set $U \subset \Lambda$ whose measure exceeds $1 - \delta'^2$. For any fixed m

$$\mu(U) = \int_{\Lambda} [\text{Freq. of oracles on } |_0^{m \cdot h(n)}](\lambda) d\lambda.$$

Hence an application of Chebyshev allows us to conclude (8).

To verify (9), apply the ergodic theorem to \mathcal{B} . Pick n so large that most \mathcal{B} -orbits of length $h(n)$ or longer visit \mathbf{H} with about the correct frequency. So this certainly holds for orbits of length $m \cdot h(n)$. □

(7.1.2) *Definition.* For any fixed implicit ζ we say a ζ -index w is *oracular with respect to β* if there is an $i \in \mathbb{N}$ such that $\text{Weight}(\beta|_0^i) = w$ and ${}_n\mathbf{A}^{\mathcal{S}^i\langle\beta\rangle} = \mathring{\mathbf{A}}(\mathcal{X}^w\zeta)$; in other words, such that $\mathcal{S}^i\langle\beta\rangle$ equals $\langle\mathcal{X}^w\beta\rangle$ and is an oracle-pair.

(7.1.3) **LEMMA.** $\forall \delta, \forall$ ‘high’, $\forall_{\text{large } n}, \forall m$ there is a set $U \subset \Lambda$ with $\mu(U) > 1 - \delta$, such that for $\langle\beta\rangle \in U$:

$$\text{A ‘high’ frequency of } w \in [0, W) \text{ are oracular with respect to } \beta, \tag{10}$$

where $W \triangleq \text{Weight}(\beta|_0^{m \cdot h(n)})$.

Proof. Fix δ and force n large enough that the previous lemma gives us sets U' and V sufficiently large that $U \triangleq U' \cap V$ has enough measure. Pick some $\langle\beta\rangle \in U$. The frequency of $i \in [0, mh(n))$ such that $\mathcal{S}^i\langle\beta\rangle$ is an oracle-pair, exceeds $1 - \delta'$; this δ' , of the previous lemma, may have been chosen as small as desired. Let $G \triangleq \{i \in [0, mh(n)): \beta_i = \mathbf{H}\}$. The density of G in $[0, mh(n))$ exceeds the fixed positive number $\mu(\mathbf{H})/2$. Consequently, we can have chosen δ' so small that a high frequency of $i \in G$ are such that $\mathcal{S}^i\langle\beta\rangle$ is an oracle-pair. This implies (10) in light of the observation that $i \mapsto \text{Weight}(\beta|_0^i)$ is a bijection from G to $[0, W)$. \square

It is convenient to restate this lemma by applying Chebyshev to U .

(7.1.3') **COROLLARY.** \forall ‘most’ and ‘high’, $\forall_{\text{large } n}, \forall m$ most ζ are such that for most β : a high frequency of $w \in [0, W)$ are oracular with respect to β and are such that

$$w \in I_k \Rightarrow ({}_n\mathbf{A}\langle\beta\rangle)|_k = \mathring{\mathbf{A}}(\mathcal{X}^w\zeta).$$

Our eventual contradiction will arise from comparing the parsing intervals $\{I_k\}_k$ and $\{J_l\}_l$ of two points $\langle\beta\rangle, \langle\rho\rangle \in \Lambda$.

(7.1.4) **SIMULTANEOUS ORACLE THEOREM.** \forall ‘most’ and ‘high’, $\forall_{\text{large } n}, \forall m$ most ζ are such that most couples $\langle\beta, \rho\rangle \in \Omega_{\mathcal{B}} \times \Omega_{\mathcal{B}}$ satisfy: a high frequency of values w in the interval $w \in [0, \text{Weight}(\beta|_0^{m \cdot h(n)}))$ are ‘simultaneously oracular’ i.e. are such that

$$w \in I_k \cap J_l \Rightarrow ({}_n\mathbf{A}\langle\beta\rangle)|_k = \mathring{\mathbf{A}}(\mathcal{X}^w\zeta) = ({}_n\mathbf{A}\langle\rho\rangle)|_l.$$

Proof. By the weak law of large numbers $\text{Weight}(\beta|_0^{m \cdot h(n)})$ is, up to a small percentage difference, equal to $\text{Weight}(\rho|_0^{m \cdot h(n)})$ for most couples $\langle\beta, \rho\rangle$. In light of this, the theorem follows from (7.1.3') since a set of measure arbitrarily close to 1, intersected with another such set, is itself such a set. Similarly, the intersection of a high density sequence of integers with another such sequence, itself has high density. \square

(7.2) *A Bernoulli randomness result.* (The proof herein was substantially simplified due to conversations with Benjamin Weiss and Steven Kalikow.) This subsection is devoted to proving an intuitive idea about an independent process \mathcal{B} , about coin flipping. So, for the duration of this subsection, we can repress any knowledge connected with \mathcal{T}, \mathcal{S} , or \mathcal{X} - and so there should be no confusion if we write $\Omega_{\mathcal{B}}$ as Ω . The product measure on $\Omega \times \Omega$ will be denoted unimaginatively by $\mu^2(\cdot)$. All the results below are stated in terms of blockings of length h . When we later use these results to get a contradiction, the role of h will be played by $h(n)$.

We study sequences of integers $\{w_k\}_{k=0}^\infty$, called *hashmarks*, usually obtained from some Bernoulli name β via

$$w_k = w_k(\beta) \triangleq \text{Weight}(\beta|_0^{k \cdot h}).$$

For such a sequence, the expected length of $w_{k+1} - w_k$ is $h \cdot \mu(\mathbf{H})$; we henceforth denote this quantity by \mathcal{E} .

We will compare sequences of hashmarks $\{w_k(\beta)\}_{k=0}^\infty$ and $\{w_l(\rho)\}_{l=0}^\infty$ obtained from two Bernoulli names $\beta, \rho \in \Omega$. Unless otherwise indicated, w_k will abbreviate $w_k(\beta)$ and W_l abbreviates $w_l(\rho)$. Our goal is to show that if we pick β and ρ at random, the sequences $\{w_k\}_{k=0}^\infty$ and $\{W_l\}_{l=0}^\infty$ are frequently ‘out of phase’ in the sense that when an interval of integers $[w_k, w_{k+1})$ intersects some $[W_b, W_{b+1})$ then the length of their intersection usually exceeds some reasonable percentage of \mathcal{E} . I chose, pretty much arbitrarily, $(1/117)\mathcal{E}$ to be a ‘reasonable percentage’.

Let α be some small number – say, $\alpha \triangleq 1/2^{100}$. We use ‘ \approx ’ as follows: $a \approx b$ means the number a is within α -percent of the number b .

(7.2.1) *Preliminaries.* Fix a specific independent process by fixing a value $\mu(\mathbf{H}) \in (0, 1)$. The weak law of large numbers tells us that

$$\mu \left\{ \beta: \frac{1}{h} |\text{Weight}(\beta|_0^h) - \mathcal{E}| < \alpha \right\} \rightarrow 1, \quad \text{as } h \rightarrow \infty. \tag{11}$$

Hence, we can fix an h sufficiently large that, by the ergodic theorem, (almost) every point $\beta \in \Omega$ satisfies:

$$w_{k+1} - w_k \triangleq \text{Weight}(\beta|_{kh}^{(k+1)h}) \approx \mathcal{E} \quad \text{for all but } (\alpha/100)\text{-percent of } k \in \mathbb{N}_0. \tag{12}$$

In the sequel, we treat \mathcal{E} – the expected number of Heads in an h -word, as a large integer; for we can have chosen h sufficiently large that \mathcal{E} is an integer to within a small percentage error – say, to within α -percent.

(7.2.2) *Definition.* Given a non-decreasing sequence of integers $\{w_k\}_{k=0}^\infty$ we say that an index k is *typical* if $w_{k+1} - w_k \approx \mathcal{E}$. For any $j \in \mathbb{N}$, say that an index k is *j-typical* if the indices $k-j, k-j+1, \dots, k+j$ are each typical. In this terminology, (12) implies that for all β : k is 3-typical of $\{w_k(\beta)\}_{k=0}^\infty$ for all but $(\alpha/3)$ -percent of $k \in \mathbb{N}_0$.

Given another non-decreasing sequence $\{W_l\}_{l=0}^\infty$ we can define a ‘correspondence function’ $l: \mathbb{N} \rightarrow \mathbb{N}$ by $l(k)$ is the smallest typical (of the second sequence) value $l \geq 0$ such that $W_l - w_k \geq 0$.

In order to be consistent with our $|_0^M$ notation we shall, given any M in the set $\{\infty, 1, 2, \dots\}$, interpret a symbol like $\{w_k\}_{k=0}^M$ to mean $\{w_k: k \in |_0^M\}$. We say that a finite or infinite sequence $\{w_k\}_{k=0}^M$ is $X\%$ out of phase with respect to $\{W_l\}_{l=0}^\infty$ if, for at least $X\%$ of $k \in |_0^M$,

$$|I_k \cap J_{l(k)}|, |I_{k+1} \cap J_{l(k)}| > \frac{1}{117} \mathcal{E}. \tag{13}$$

Here, for general k and l , I_k abbreviates the interval of integers $[w_k, w_{k+1})$ and J_l denotes $[W_l, W_{l+1})$.

(7.2.3) *LEMMA.* For any $(\beta, \rho) \in \Omega \times \Omega$ at least $(1 - 4\alpha)$ -percent of $k \in \mathbb{N}_0$ are such that k is 3-typical,

$$l(k) \text{ is 1-typical, and } W_{l(k)} - w_k < (1 + \alpha)\mathcal{E}. \tag{14}$$

Proof. Define $l'(k)$ to be the smallest 3-typical $l \geq 0$ such that $W_l \geq w_k$. It suffices to establish the lemma with (14) replaced by

$$W_{l'(k)} - w_k < 2\mathcal{E}. \tag{14'}$$

Say that a 3-typical value, k , colours an index $i \in \mathbb{N}$ *blue* if $w_k \leq i < w_k + (1 - \alpha)\mathcal{E}$. Since the righthand side is less than w_{k+1} , an index i can be coloured blue by at most one k . Hence, by the strong law of large numbers, the set of blue i has (lower) density exceeding $(1 - (\alpha/3))(1 - \alpha) \stackrel{\text{note}}{>} 1 - 2\alpha$. Similarly, the set of *yellow* indices – those i such that $W_l - (1 - \alpha)\mathcal{E} < i \leq W_l$ for some 3-typical l , also has density greater than $1 - 2\alpha$. So the set – call it G – of i which are both blue and yellow, has density exceeding $1 - 4\alpha$. Thus, the set of values k which colour blue some $i \in G$, has density dominated by $(1 - 4\alpha)/(1 - \alpha)$ and hence by $1 - 4\alpha$. Such k satisfy (14'). \square

(7.2.4) WEAK OUT-OF-PHASE LEMMA. $\exists M_0$ such that $\forall M \geq M_0$ and for each ‘shift value’ $q \in \{0, 1, \dots, h\}$, there is a set $V_q \subset \Omega \times \Omega$ with $\mu^2(V_q) > 7/10$, such that whenever $(\beta, \rho) \in V_q$ the sequence $\{q + w_k\}_{k=0}^M$ is 50% out of phase with respect to $\{W_l\}_{l=0}^\infty$.

Proof. Since there are but a fixed number of values q to consider, it suffices to prove the result with M_0 allowed to depend on q . As will be clear from the proof, we may assume without essential loss of generality, that $q = 0$. So, it suffices to prove that for each fixed $\rho \in \Omega$ the following holds. $\forall_{\text{large}} M$ and for at least measure 8/10 of $\beta \in \Omega$ at least 50% of $k \in \{0, \dots, M\}$ are scattered. A k is scattered by β if: k and $k + 1$ are typical; $l = l(k)$ is typical; the differences $W_{l+1} - w_{k+1}$ and $w_{k+1} - W_l$ each exceed $(1/117)\mathcal{E}$. We proceed to the demonstration of the above, leaving M to be specified.

Define $\text{freq}(\beta)$ to be the density of $k \in \{0, \dots, M\}$ which are scattered by β . Say that β is good if $\text{freq}(\beta) > 50\%$. We construct a bi-measure preserving map $f_1: \Omega \rightarrow \Omega$. Define $k_0(\beta)$ to be the smallest $k > 0$ such that

$$\beta|_{kh}^{(k+1)h} = \underbrace{\text{T T T} \dots \text{T}}_h.$$

Define $k_1(\beta)$ to be the most positive $k < 0$ such that

$$\beta|_{kh}^{(k+1)h} = \underbrace{\text{H T T} \dots \text{T}}_h.$$

We now let f_1 be the map which ‘switches the h -words at k_0 and k_1 ’ i.e.

$$f_1(\beta)|_{kh}^{(k+1)h} \triangleq \begin{cases} \text{H T T} \dots \text{T} & \text{if } k = k_0(\beta); \\ \text{T T T} \dots \text{T} & \text{if } k = k_1(\beta); \\ \beta|_{kh}^{(k+1)h} & \text{otherwise.} \end{cases}$$

Notice that $\forall k > k_0(\beta)$: $\text{Weight}(f_1(\beta)|_0^{kh}) = 1 + \text{Weight}(\beta|_0^{kh})$, which is to say that $w_k(f_1(\beta)) = 1 + w_k$.

For each $i \in \{1, \dots, \mathcal{E}\}$ we can, in this fashion, define a bi-measure preserving function $f_i: \Omega \rightarrow \Omega$ and achieve that for each $\beta \in \Omega$:

$$\forall_{\text{large}} k \text{ and } \forall i \in \{1, \dots, \mathcal{E}\}: w_k(f_i(\beta)) = i + w_k. \tag{15}$$

At this juncture, fix some particular β and let $K = K_\beta$ be the set of k large enough for (15), such that $k, k + 1, l(k), l(k) + 1$ are typical, and for which

$W_{l(k)} - w_k < (1 + \alpha)\mathcal{E}$. By the preceding lemma, the (lower) density of K in \mathbb{N} exceeds $1 - 4\alpha$. For a fixed $k \in K$, notice that for at least $(1 - 100\alpha - \frac{2}{117})$ -percent of subscripts $i \in \{1, \dots, \mathcal{E}\}$:

$$(1 - \alpha - \frac{1}{117})\mathcal{E} > W_{l(k)} - (i + w_k) > (2\alpha + \frac{1}{117})\mathcal{E},$$

where $l_i(k)$ is to mean $l(k)$, if $W_{l(k)} \geq i + w_k$, and $l(k) + 1$ otherwise. If \mathcal{E} were truly an integer the expression would be $(1 - 4\alpha - \frac{2}{117})$ -percent; since \mathcal{E} may be as far as α -percent from being an integer, we take $(1 - 100\alpha - \frac{2}{117})$ to be safe. Consequently the differences $W_{l(k)+1} - (i + w_{k+1})$ and $(i + w_{k+1}) - W_{l(k)}$ each exceed $(\frac{1}{117})\mathcal{E}$. This says that k is scattered by $f_i(\beta)$.

Now fix some M sufficiently large that, for most β , the set of indices K_β has density in $|_0^M$ exceeding $1 - 5\alpha$; in fact, no harm will come to us if we pretend this is true for all β . Setting $1 - \varepsilon \triangleq (1 - 100\alpha - \frac{2}{117})$ and letting k range over $|_0^M$, we get, by a Chebyshev argument, that for $(1 - \varepsilon)$ -percent of $i \in \{1, \dots, \mathcal{E}\}$ that the number $\text{freq}(f_i(\beta))$ exceeds $(1 - \varepsilon)$ times the density of the set K_β in $|_0^M$. Hence this number exceeds $(1 - \varepsilon)(1 - 5\alpha) \stackrel{\text{note}}{>} 50\%$. This tells us, for each $\beta \in \Omega$, that for at least $(1 - \varepsilon)$ -percent of $i \in \{1, \dots, \mathcal{E}\}$, the point $f_i(\beta)$ is good. Since each f_i is measure preserving, the measure of the set of good β exceeds $1 - \varepsilon$. This certainly exceeds $8/10$. □

The preceding lemma is ‘weak’ since a set V_q may well depend on M . With the strong law of large numbers and a straightforward stopping time argument, we can convert it into the following a.e. statement.

(7.2.5) **BOOTSTRAP LEMMA.** *For any $\langle \beta, \rho \rangle \in \Omega \times \Omega$, $\{w_k\}_{k=0}^\infty$ is at least 12% out of phase with respect to $\{W_i\}_{i=0}^\infty$.*

Sketch of proof. This lemma makes an assertion about a pair of randomly chosen half-infinite sequences of Heads and Tails, $\beta|_0^\infty$ and $\rho|_0^\infty$. We shall pick such a pair of sequences, randomly, in stages $s = 0, 1, 2, \dots$. Fix some M from (7.2.4). At stage s , we have defined $\beta|_0^{sM \cdot h}$ and $\rho|_0^{m_s \cdot h}$ where m_s will have been chosen so that the number

$$q_s \triangleq \text{Weight}(\rho|_0^{m_s \cdot h}) - \text{Weight}(\beta|_0^{sM \cdot h}) \tag{16}$$

is in $|_0^h$. To pass to stage $s + 1$ we pick, independently of all previous choices, a point $\langle \gamma, \varrho \rangle \in \Omega \times \Omega$. Set

$$\beta|_{sMh}^{sMh + Mh} \triangleq \gamma|_0^{Mh} \quad \text{and} \quad \rho|_{m_s h}^{m_s h + m' h} \triangleq \varrho|_0^{m' h}$$

and set $m_{s+1} \triangleq m_s + m'$, where $m' \geq 0$ is chosen smallest such that q_{s+1} , defined as in (16), is non-negative.

Denote the couple $\langle \gamma, \varrho \rangle$ picked at stage s by ω_s . For any particular $q \in |_0^h$, the point ω_s has probability exceeding $7/10$ of being in the set V_q of (7.2.4). Since $\omega_1, \omega_2, \dots$ are plucked from $\Omega \times \Omega$ independently of each other, the strong law of large numbers assures, with probability one, that frequency exceeding $7/10$ of $s \in |_0^\infty$ are such that $\omega_s \in V_q$. And $(7/10) \cdot 50\%$ exceeds 12% . □

(7.2.6) **COROLLARY.** *For all $\langle \beta, \rho \rangle \in \Omega \times \Omega$: For each $i, j \in \mathbb{N}$, the sequence $\{\text{Weight}(\beta|_0^{i+kh})\}_{k=0}^\infty$ is 12% out of phase with respect to $\{\text{Weight}(\rho|_0^{j+th})\}_{t=0}^\infty$.*

Proof. Set $i' \triangleq \text{Weight}(\beta|_0^i)$, $j' \triangleq \text{Weight}(\rho|_0^j)$, $\beta' \triangleq \mathcal{B}^i(\beta)$, and $\rho' \triangleq \mathcal{B}^j(\rho)$. We need to show

$$\{i' + w_k(\beta')\}_{k=0}^\infty \text{ is 12\% out of phase with respect to } \{j' + w_l(\rho')\}_{l=0}^\infty.$$

For each pair i' and j' , the proof of the lemma shows this true for a.e. point $\langle \beta', \rho' \rangle$ in $\Omega \times \Omega$. So, for fixed i and j , the corollary holds for a.e. $\langle \beta, \rho \rangle$ since $\mathcal{B}^i \times \mathcal{B}^j$ is measure preserving on $\Omega \times \Omega$. By discarding countably many sets of μ^2 -measure zero, one for each pair $i, j \in \mathbb{N}$, we complete the proof. \square

Up until now, all of the lemmas have concerned a specific independent process determined by a number $\mu(\mathbf{H})$. The foregoing results hold for h sufficiently large to satisfy the two requirements in the preliminaries of (7.2.1) - namely that

$$\mu \left\{ \beta : \frac{1}{h} |\text{Weight}(\beta|_0^h) - \mathcal{E}| \geq \alpha \right\} < \frac{\alpha}{100}, \tag{17}$$

and that \mathcal{E} is within α -percent of being a large integer.

It turns out that, given any interval $[a, b] \subset (0, 1)$, we can find an h_0 such that for any $h > h_0$ and any process $\mu(\mathbf{H}) \in [a, b]$, both requirements are fulfilled. As one can see from the standard Chebyshev proof of the weak law of large numbers (see [1, p. 5]), the measure of the set in (17) is dominated by $(1/\alpha^2)(1/h)\mu(\mathbf{H})[1 - \mu(\mathbf{H})]$ and hence by $(1/\alpha^2)(1/h)$. As for the second requirement, we need just choose h_0 so large that $a \cdot h_0$ exceeds, say, $100/\alpha$.

These considerations allow us to write the above corollary as follows.

(7.2.7) LEMMA. *For each interval $[a, b] \subset (0, 1)$ there exists $h_0 = h_0([a, b])$ such that for any independent process $\mu(\mathbf{H}) \in [a, b]$, any $h > h_0$, and any $\delta' > 0$, the following holds: $\forall_{\text{large } M}$ there is a set $U \subset \Omega \times \Omega$, where U is a finite union of cylinder sets and $\mu^2(U) > 1 - \delta'$, such that the following is fulfilled by all $\langle \beta, \rho \rangle \in U$. For each $i, j \in \mathbb{N}$:*

$$\{\text{Weight}(\beta|_0^{i+kh})\}_{k=0}^M \text{ is 11\% out-of-phase with respect to } \{\text{Weight}(\rho|_0^{j+lh})\}_{l=0}^\infty.$$

Proof. Immediate. That U can be taken to be a union of $|_0^{M \cdot h}$ cylinders follows from the weak law of large numbers. It tells us that for most β and ρ , we can tell whether $\langle \beta, \rho \rangle$ satisfies the above condition by looking at $\beta|_0^{M \cdot h}$ and $\rho|_0^{M \cdot h}$. \square

(7.2.8) BERNOULLI OUT-OF-PHASE THEOREM. *For each compact interval $[a, b] \subset (0, 1)$, $\forall_{\text{large } h}, \forall \delta$, there is an M_0 such that $\forall M \geq M_0$: If \mathcal{B} is an independent process for which $\mu(\mathbf{H}) \in [a, b]$, then $\mu^2(U) > 1 - \delta$ where U is the set of all $\langle \beta, \rho \rangle$ such that for each $i, j \in \mathbb{N}$:*

$$\{\text{Weight}(\beta|_0^{i+kh})\}_{k=0}^M \text{ is 10\% out-of-phase with respect to } \{\text{Weight}(\rho|_0^{j+lh})\}_{l=0}^\infty. \tag{18}$$

Remark. The improvement over lemma (7.2.7) is that $\mu(\mathbf{H})$ is quantified *after* M_0 . Thus, we ‘decouple’ our choice of M from the value of $\mu(\mathbf{H})$.

In addition, we obtain out-of-phasesness without needing to know the exact values of i and j . We are able to decouple our choice of M from having to have prior knowledge, at assemblytime, of the function $\lambda \mapsto i_0(\lambda)$. This allows us to bypass the difficulty discussed at the beginning of § 7.

Proof. Fix any h greater than the $h_0([a, b])$ of (7.2.7). Fix a δ . Just for the duration of this proof, define a function $M[\cdot]$ as follows: for any number $\bar{\mu} \in [a, b]$, let $M[\bar{\mu}]$

be the smallest value of M' which works in the preceding lemma when $\mu(\mathbf{H}) \triangleq \bar{\mu}$ and $\delta' \triangleq \delta$. Our theorem is proved should we demonstrate that $M[\cdot]$ is bounded above on $[a, b]$. For if (7.2.7) works up to 11% when M' is set equal to some number M , then it works up to 10% for multiples of M and consequently for all M' bigger than, say, $2^{100}M$. This follows by an argument like that of the bootstrap lemma along with assuming δ' sufficiently small.

By the compactness of $[a, b]$, it suffices to show that $M[\cdot]$ is bounded on some neighbourhood of any given $\bar{\mu}_0 \in [a, b]$. Let L be a value of M' which works in (7.2.7) when $\mu(\mathbf{H}) \triangleq \bar{\mu}_0$ and $\delta' \triangleq \delta/2$. That lemma hands us a finite union of cylinder sets $U \subset \Omega \times \Omega$ such that $\mu^2(U) > 1 - \delta/2$. On U , property (18) is fulfilled with $M \triangleq L$. Since U is a finite union of cylinders, its measure varies continuously as a function of the probability of Heads, $\mu(\mathbf{H})$. Consequently, if $\mu(\mathbf{H}) = \bar{\mu}$ for a $\bar{\mu}$ in a sufficiently small neighbourhood about $\bar{\mu}_0$, then

$$\mu^2(U) > (1 - \delta/2) - \delta/2 = 1 - \delta.$$

Thus, the number L bounds $M[\cdot]$ on this neighbourhood. □

Remark. So as to spare the reader the distraction of a really minor detail, we refrained from mentioning that the points $\langle \beta, \rho \rangle$ of U , in addition to fulfilling (18), can be asked to satisfy

$$\mathcal{E} > \frac{117}{118} \frac{1}{M} \text{Weight}(\beta|_0^{Mh}). \tag{19}$$

For a specific value of $\mu(\mathbf{H})$, this follows from the weak law of large numbers. The compactness argument persists since $\mathcal{E} \triangleq h \cdot \mu(\mathbf{H})$ varies continuously as a function of $\mu(\mathbf{H})$.

(7.3) *Obtaining the contradiction.* We have all the ingredients necessary to break any potential isomorphism. In order to avoid logical errors, we relist our ingredients in the order necessary for a formal proof.

We use the above theorem to choose a sequence of positive δ_h which decreases to zero as $h \rightarrow \infty$, and a function $M(\cdot)$, so that for each h the out-of-phase theorem works when δ is set equal to δ_h , $M = M(h)$, and $[a, b] = [\delta_h, 1 - \delta_h]$. This function $M(\cdot)$ is defined solely in terms of coin flipping. Hence, we can hand this $M(\cdot)$ to assembly, (3.3), as the ‘lower bound’ function. Assembly then churns out a transformation \mathcal{T} and two sequences of integers $\{h(n)\}_{n=1}^\infty$ and $\{nM\}_{n=1}^\infty$, each $nM = M(h(n))$, for which the relevant properties of (3.3.5) hold.

Next, assume that there is an isomorphism φ to \mathcal{T} from a simple independent skew $\mathcal{S} = \mathcal{B} \otimes \langle \mathcal{X}/\mathbf{H}, \text{Id}/\mathbf{T} \rangle$. We now derive the simultaneous oracle theorem using the \bar{d}_e property of (3.3.5). Also, $\mu(\mathbf{H})$ is now some specific number in $(0, 1)$ and hence, $\forall_{\text{large } n}, \mu(\mathbf{H}) \in [\delta_h, 1 - \delta_h]$ where we are letting h mean $h(n)$. Consequently, the conclusion of the out-of-phase theorem holds for a set of $\langle \beta, \rho \rangle$ of measure exceeding δ_h . Since the δ_h decrease to zero as h , hence n , goes to infinity, we can restate the theorem thus:

(7.3.1) *Out-of-phase revisited.* \forall ‘most’, $\forall_{\text{large } n}$, most $\langle \beta, \rho \rangle$ satisfy: for each $i, j \in |_0^{h(n)}$, $\{\text{Weight}(\beta|_0^{i+kh(n)})\}_{k=0}^{nM}$ is 10% out-of-phase with respect to $\{\text{Weight}(\rho|_0^{j+lh(n)})\}_{l=0}^{nM}$.

Also,

$$\frac{117}{118} \frac{1}{nM} \text{Weight}(\beta|_0^{n^{M \cdot h(n)}}) < \mathcal{E}.$$

As our next step, we can lift the infrequency of duplications property (3.3.5(c)) to the space Λ via φ . An application of Chebyshev allows it thus:

(7.3.2) *Infrequency-of-duplications revisited.* \forall ‘most’, \forall large n , for most ζ we find that most β are such that the parsing $\{I_k\}_{k=0}^\infty$ of ζ induced by β , satisfies the following:

- (i) For most $k \in [0, nM)$: $(nA\langle\beta\rangle)|_k \neq (nA\langle\beta\rangle)|_{k+1}$.
- (ii) The intervals $\{I_k\}_0^{nM}$ are contiguous and the lefthand endpoint of I_0 is in $|_0^{h(n)}$.

We now combine the infrequency of duplications, simultaneous oracle, and out-of-phase theorems.

(7.3.3) **CONTRADICTION THEOREM.** \forall ‘most’, \forall large n , most ζ are such that most couples $\langle\beta, \rho\rangle$ satisfy the following: Let $\{I_k\}_{k=0}^\infty$ be the parsing of ζ induced by β . Let $\{J_l\}_{l=0}^\infty$ be the parsing of ζ by ρ . Let W be the number $\text{Weight}(\beta|_0^{n^{M \cdot h(n)}})$. Then

- (i) For most $k \in |_0^{nM}$: $(nA\langle\beta\rangle)|_k \neq (nA\langle\beta\rangle)|_{k+1}$.
- (ii) The intervals $\{I_k\}_0^{nM}$ are contiguous and the lefthand endpoint of I_0 is in $|_0^{h(n)}$. The same holds for $\{J_l\}_{l=0}^{nM}$.
- (iii) Most $w \in [0, W)$ are simultaneously oracular i.e. are such that:

$$w \in I_k \cap J_l \Rightarrow (nA\langle\beta\rangle)|_k = (nA\langle\rho\rangle)|_l.$$

- (iv) Over 10% of $k \in |_0^{nM}$ are such that:

$$|I_k \cap J_{l(k)}|, |I_{k+1} \cap J_{l(k)}| > \frac{1}{117} \mathcal{E} > \frac{1}{118} \frac{1}{nM} W.$$

- (v) For at least 8% of $k \in |_0^{nM}$: $(nA\langle\beta\rangle)|_k = (nA\langle\beta\rangle)|_{k+1}$.
- (vi) *Contradiction.*

Proof. The first two properties come from (7.3.2) applied to both $\langle\beta\rangle$ and $\langle\rho\rangle$. The next, (iii), is a restatement of the simultaneous oracle theorem (7.1.4).

We will justify (iv). First note that since the $\{I_k\}_0^{nM}$ are contiguous, their endpoints form a sequence of hashmarks just as we defined them. Saying that these endpoints are ‘out-of-phase’ (defined in (13)) is what the first inequality of (iv) asserts. The second inequality follows because

$$\frac{1}{117} \mathcal{E} > \frac{1}{117} \cdot \frac{117}{118} \frac{1}{nM} \text{Weight}(\beta|_0^{n^{M \cdot h(n)}}) = \frac{1}{118} \frac{1}{nM} W.$$

Now to verify (v). We call any index k satisfying (iv), *good*. As k ranges over good values, the intervals $I_k \cap J_{l(k)}$ are all disjoint. Consequently the union of these intervals is a fixed percentage of $[0, W)$. Specifically,

$$\frac{1}{W} \cdot \left| \bigcup_{k \text{ good}} I_k \cap J_{l(k)} \right| > 10\% \cdot \frac{1}{118} = \frac{1}{10} \cdot \frac{1}{118}.$$

We can have chosen n large enough that the frequency of simultaneously oracular $w \in [0, W)$ is so high that at least 9% of $k \in |_0^{nM}$ are at once good *and* such that $I_k \cap J_{l(k)}$ contains a simultaneously oracular w ; hence such that

$({}_n\mathbf{A}\langle \xi/\beta \rangle)|_k = ({}_n\mathbf{A}\langle \xi/\rho \rangle)|_{I(k)}$. Repeating this argument for $I_{k+1} \cap J_{I(k)}$ in place of $I_k \cap J_{I(k)}$ allows us to conclude (v).

Finally, (v) and (i) add up to (vi). □

8. Results

Here is what we have shown:

There is a \mathcal{K} -automorphism \mathcal{T} not isomorphic to any simple independent skew.

For simplicity, in our proof we tacitly assumed that \mathcal{L} 's generating partition \mathbf{P} was finite, thus forcing \mathcal{L} to have finite entropy. Also, we assumed that φ was an isomorphism:

$$\begin{array}{ccc} \Lambda & \xrightarrow{\mathcal{S}} & \Lambda \\ \downarrow \varphi & & \downarrow \varphi \\ X & \xrightarrow{\mathcal{T}} & X \end{array}$$

But if we allow \mathbf{P} to be countable, and φ to be just a homomorphism, then our proof goes through without so much as a notational change. Hence we may conclude: *There exists a \mathcal{K} -automorphism \mathcal{T} which is not a factor of any ergodic transformation \mathcal{S} of the form $\mathcal{B} \otimes \langle \mathcal{L}/\mathbf{H}, \text{Id}/\mathbf{T} \rangle$ where \mathcal{L} is any transformation (of zero, positive, or infinite entropy) and where \mathcal{B} is Bernoulli with independent generator $\langle \mathbf{H}, \mathbf{T} \rangle$ such that the skewing set obeys $0 \leq \mu(\mathbf{H}) < 1$.*

Proof. We have shown this for $0 < \mu(\mathbf{H}) < 1$. The $\mu(\mathbf{H}) = 0$ case follows by letting \mathcal{L} be the identity transformation on a one point space $\Omega_{\mathcal{L}}$. □

Remark. In the case $\mu(\mathbf{H}) = 0$, the skew product \mathcal{S} is the Bernoulli base transformation \mathcal{B} . Hence our \mathcal{K} -automorphism \mathcal{T} is not Bernoulli.

With regard to the direct product case, where $\mu(\mathbf{H}) = 1$, not only does the proof break down, but so does the result. Recall that we built \mathcal{T} by concatenating n -letters. As a consequence of the independent way that our members are chosen from n -letters (see the description of grouping following figure 2) the Thouvenot theory implies that T satisfies the weak Pinsker conjecture [11]. We remark without proof that \mathcal{T} is a counterexample to the Pinsker conjecture, that is, when \mathcal{L} is constrained to have zero entropy.

Remark. Observe that the only place we used that the partition $\langle \mathbf{H}, \mathbf{T} \rangle$ generated under \mathcal{B} was in § 6 to manufacture oracles. With notational changes we can make the proof of § 6 go through with the partition $\langle \mathbf{H}, \mathbf{T} \rangle$, rather than generating, just required to be refined by some independent generating (possibly countable) partition. We obtain the following result.

THEOREM. *There exists a \mathcal{K} -automorphism \mathcal{T} which is not a factor of any ergodic transformation \mathcal{S} of the form $\mathcal{B} \otimes \langle \mathcal{L}/\mathbf{H}, \text{Id}/\mathbf{T} \rangle$ where \mathcal{L} is any transformation (of zero, positive, or infinite entropy), where \mathcal{B} is any (finite or infinite entropy) Bernoulli transformation with a countable independent generating partition that refines $\langle \mathbf{H}, \mathbf{T} \rangle$, and where the skewing set obeys $0 \leq \mu(\mathbf{H}) < 1$.*

Actually, the proof in this paper shows that, in the space of processes, there is a small \bar{d} -ball (a radius of $(1/10)\text{Gap}$ will work) centred about the \mathcal{T} -process such that no transformation in this ball is a factor of any simple independent skew.

This work was done for the author's Ph.D. dissertation written at Stanford University under the direction of D. S. Ornstein.

REFERENCES

- [1] L. Breiman. *Probability*. Addison-Wesley, 1968.
- [2] R. Burton. A non-Bernoulli skew product which is loosely Bernoulli. *Israel J. Math.* **35** (1980), no. 4, 339–348.
- [3] N. Friedman. *Introduction to Ergodic Theory*. Van Nostrand Reinhold Mathematical Studies #29, 1970.
- [4] I. Meilijson. Mixing properties of a class of skew products. *Israel J. Math.* **19** (1974), 266–270.
- [5] D. S. Ornstein. A mixing transformation for which Pinsker's conjecture fails. *Adv. in Math.* **10**, No. 1, Feb. 1973, p. 103.
- [6] D. S. Ornstein & P. Shields. An uncountable family of K -automorphisms. *Adv. in Math.* **10**, No. 1, Feb. 1973, p. 63.
- [7] D. Ornstein, D. Rudolph & B. Weiss. *Equivalence of measure preserving transformations*. *Mem. Amer. Math. Soc.* **37**, No. 262, May 1982.
- [8] D. Rudolph. Two non-isomorphic K -automorphisms with isomorphic squares. *Israel J. Math.* **23** (1976), 274–287.
- [9] P. Shields. *The Theory of Bernoulli Shifts*. University of Chicago Press: Chicago and London, 1973.
- [10] P. Shields & R. Burton. A skew product which is Bernoulli. *Mh. Math.* **86** (1978), 155–165.
- [11] Jean-Paul Thouvenot. Quelques propriétés des systèmes dynamiques qui se décomposent en un produit de deux systèmes dont l'un est un schéma de Bernoulli. *Israel J. Math.* **21** (1975), 177–207.