

# The commutant is the weak closure of the powers, for rank-1 transformations

JONATHAN KING

Department of Mathematics and Statistics, State University of New York at Albany,  
Albany, NY 12222, USA

(Received 13 March 1985; revised 1 July 1985 and 10 October 1985)

*Abstract.* In the class of rank-1 transformations, there is a strong dichotomy. For such a  $T$ , the commutant is either *trivial*, consisting only of the powers of  $T$ , or is *uncountable*. In addition, the commutant semigroup,  $C(T)$ , is in fact a group. As a consequence, the notion of weak isomorphism between two transformations is equivalent to isomorphism, if at least one of the transformations is rank-1. In § 2, we show that any proper factor of a rank-1 must be rigid. Hence, neither Ornstein's rank-1 mixing nor Chacón's transformation, can be a factor of a rank-1.

N.B. In this paper 'transformation' shall mean a measure preserving transformation on a Lebesgue probability space. In general, we assume our measure spaces non-atomic; however the spaces that factors live on may be atomic. A transformation  $T$  is *invertible* if  $T^{-1}$  exists and is a transformation. All sets mentioned are assumed measurable.

0. Suppose  $R: [0, 1) \rightarrow [0, 1)$  is an irrational rotation: the transformation sending  $x \mapsto x + \alpha \pmod{1}$ , for some fixed irrational number  $\alpha$ . Now let  $S$  be rotation by some amount  $\beta$ .  $S$  commutes with  $R$ , i.e.  $RS = SR$ . This  $S$  can be obtained as a certain limit of powers of  $R$ . By the irrationality of  $\alpha$ , there exists a sequence of integers  $s_i \rightarrow \infty$  such that  $s_i \cdot \alpha \pmod{1} \rightarrow \beta$  as  $i \rightarrow \infty$ . Thus  $R^{s_i} \rightarrow S$  weakly as  $i \rightarrow \infty$ . In general, for transformations  $S, T_i: X \rightarrow X$  we say that  $T_i \rightarrow S$  weakly if, for all  $A \subset X$ ,

$$\mu(T_i^{-1}(A) \Delta S^{-1}(A)) \rightarrow 0 \quad \text{as } i \rightarrow \infty.$$

This is equivalent to saying that the operators on  $L^2(X)$  associated with the  $T_i$  converge weakly (i.e. on each  $f \in L^2$ ) to the operator associated with  $S$ . The limit,  $S$ , need not be invertible, even if all the  $T_i$  are. It is not hard to show, however, that:

(0.1) *If the  $T_i$  are invertible and they commute with each other, then  $S$  is invertible and  $T_i^{-1} \rightarrow S^{-1}$  weakly.*

In particular,  $S$  will be invertible if the  $T_i$  are various powers of a single transformation.

In [4], Andres del Junco states that for rank-1 transformations  $T$  possessing a certain special name structure, any transformation  $S$  commuting with  $T$  is a weak limit of powers of  $T$ . He remarks that it is unknown whether this holds for a general rank-1. Perhaps he was led to this question by his study of irrational rotations (in [5] he shows them to be rank-1) since one might suspect that any transformation commuting with a rotation must be, itself, a rotation.

The goal of this paper is to answer del Junco’s question in the affirmative and to derive several corollaries. Also, we construct two weak-mixing rank-1 transformations which exhibit behaviour showing that the limitations of the corollaries are not vacuous.

*Notations and conventions.* As is customary, proofs and statements have a tacit ‘a.e.’ attached. Partitions have finitely many atoms. Given a partition  $P$ , a  $P$ -name is an infinite string of letters from  $P$ . A point  $x$  in the space has a doubly infinite  $T, P$ -name

$$x = \cdots x_{-1}x_0x_1x_2 \cdots = x|_{-\infty}^{\infty}.$$

A  $P$ -word is a finite string of letters. If  $W$  is a word of length  $h$ , we index  $W$  as

$$W = W_0W_1 \cdots W_{h-1} = W|_0^h.$$

By extension we use  $|_a^b$  to denote the half-open interval of integers  $[a, b)$ .  $W|_i$  is a synonym for  $W_i$ .

We let  $\text{Id}$  denote the identity transformation on any space under discussion. Agree to let  $a \triangleq b$  or  $b \triangleq a$  to mean that the expression  $b$  defines the symbol  $a$ . This is useful for emphasis and for defining abbreviating terms in the middle of an equation. Also, a statement like ‘ $(1 - \varepsilon)/5 \triangleq 3$ ’ means to define  $\varepsilon$  so that the equation holds.

Symbols  $\varepsilon, \delta, e$  and  $\sigma$  always represent small quantities in  $(0, 1)$ . ‘ $\forall$  large  $n$ ’ means ‘ $\exists N$  such that  $\forall n \geq N$ ’. The symbol  $\square$  ends a proof.

*Definition.* Given an invertible transformation  $T: X \rightarrow X$ , we use  $C(T)$  to denote the *commutant* of  $T$ ; the set of transformations  $S: X \rightarrow X$  which commute with  $T$ .  $C(T)$  is automatically a semigroup and, should every  $S \in C(T)$  happen to be invertible, a group – in general, non-abelian. If  $C(T)$  turns out to be a group then  $\{T^n\}_{n=-\infty}^{\infty}$  is a normal subgroup and we can define the quotient group  $C(T)/\{T^n\}_{-\infty}^{\infty}$ . We call this quotient the *essential commutant* of  $T$  and denote it  $EC(T)$ . We say that the commutant is *trivial* if  $C(T) = \{T^n\}_{-\infty}^{\infty}$  i.e.  $EC(T)$  is the trivial group.

Let  $\text{WCl}(T)$  denote the weak closure of  $\{T^n\}_{-\infty}^{\infty}$ . Thus  $S \in \text{WCl}(T)$  if and only if there exists  $\{s_i\}_1^{\infty}$  such that  $S = w \lim_{i \rightarrow \infty} T^{s_i}$ ;  $\text{WCl}(T)$  is an abelian group and subset of  $C(T)$ . Let  $\text{WCl}^+(T)$  denote those  $S$  obtainable by a sequence  $s_i \rightarrow +\infty$ . We see that

$$\begin{aligned} (0.2) \quad \text{WCl}(T) \text{ non-trivial} &\Rightarrow \text{WCl}^+(T) \text{ non-empty} \\ &\Leftrightarrow \text{Id} \in \text{WCl}^+(T) \tag{0.2} \\ &\Leftrightarrow \text{WCl}^+(T) = \text{WCl}(T). \end{aligned}$$

In a moment we will see that the first implication is reversible for ergodic  $T$  (on non-atomic  $X$ ). One says an ergodic  $T$  is *rigid* if  $\text{Id} \in \text{WCl}^+(T)$ . In [7] is the handy observation (later made by del Junco, independently) that any rigid transformation

has uncountable commutant. All one needs is a sequence of non-identity transformations  $T^{s_i} \rightarrow \text{Id}$ . The group of invertible transformations on  $X$ , with the weak limit topology, becomes a metric space via the norm

$$|T| \triangleq \sum_{k=1}^{\infty} \mu(TA_k \Delta A_k)/2^k,$$

where the fixed sets  $\{A_k\}_1^{\infty}$  were chosen to be a generating algebra of sets. By dropping to a subsequence of the  $\{s_i\}$ , we may assume that for each  $i$ ,  $|T^{s_i}| > \sum_{j=i+1}^{\infty} |T^{s_j}|$ . Now for each binary vector  $b = \langle b_1 b_2 \dots \rangle$ , where  $b_i \in \{0, 1\}$ , define the transformation

$$S_b \triangleq w \lim_{i \rightarrow \infty} T^{m_i} \quad \text{where } m_i \triangleq \sum_{j=1}^i b_j \cdot s_j.$$

If  $i$  is the smallest such that  $b'_i \neq b_i$ , for binary vector  $b'$ , then  $S_{b'} \neq S_b$  since

$$|S_{b'}^{-1} S_b| \geq |T^{s_i}| - \sum_{j=i+1}^{\infty} |T^{s_j}| > 0.$$

So  $\text{WCl}(T)$ , hence  $C(T)$ , is uncountable.

*Definition.* We say that  $T$  is rank-1 if it can be built by cutting and stacking with but one column at each stage. Specifically, we can find base sets  $B_n \subset X$  and heights  $h_n \rightarrow \infty$  such that for all  $n$ , the sets  $B_n, T(B_n), \dots, T^{h_n-1}(B_n)$  are disjoint and there is some subset  $E \subset [0, h_n - h_{n-1}]$  of levels such that  $B_{n-1} = \bigcup_{i \in E} T^i(B_n)$ . Also, the sets  $\{T^i(B_n) : n \in \mathbb{N}, 0 \leq i < h_n\}$  generate. For a rank-1  $T$  one can always find a generating partition  $P$  and stacks as above where each stack has but one  $P$ -column. This  $P$ - $h_n$ -word is called the  $n$ -block and denoted  $W_{(n)}$ . Usually the  $n$  is implicit and we write  $h$  and  $W$  for  $h_n$  and  $W_{(n)}$ . Factors of rank-1 are rank-1. Rank-1 is automatically ergodic.

In the sequel,  $T : X \rightarrow X$  is a fixed rank-1 transformation (on a non-atomic space).

**DICHOTOMY THEOREM.** *The commutant of  $T$  is trivial or uncountable. In other words,  $EC(T)$  is the trivial group of one element or  $EC(T)$  is an uncountable group.*

*Proof.* This follows from the weak-closure theorem, below, by (0.2) and the result that rigid transformations have uncountable commutant. □

**WEAK-CLOSURE THEOREM.**  $\text{WCl}(T) \supset C(T)$ .

This yields several corollaries.

(A)  $C(T)$  equals  $\text{WCl}(T)$  and is therefore an abelian group. Any factor algebra of  $T$  is invariant under each  $S \in C(T)$ . Given any transformation  $F$  which is a factor of  $T$ , we get a natural (into) group homomorphism  $\gamma : C(T) \rightarrow C(F)$  simply by mapping each  $S \in C(T)$  to its restriction to the sub-sigma-algebra which determines  $F$ .  $\gamma$  need neither be injective nor surjective. However if  $T$  is non-rigid and the factor  $F$  is non-atomic, then  $\gamma$  is injective (contrast example (iv), below) and  $T$  cannot be a group extension of  $F$ .

(B) If  $T$  is weakly isomorphic to some transformation  $R: Y \rightarrow Y$  via factor maps  $\phi: X \rightarrow Y$  and  $\psi: Y \rightarrow X$ , then  $T \cong R$  and  $\phi$  and  $\psi$  are each isomorphisms.

(C)  $T$  is not a cartesian square; consequently, if  $T$  is of the form  $R_1 \times R_2$  then  $R_1$  and  $R_2$  must be relatively prime i.e. no factor of  $R_1$  is isomorphic to any factor of  $R_2$ . Moreover, if  $T$  is non-rigid,  $T$  cannot be a cartesian product. (Contrast example (vi).)

(D) For any  $k \geq 2$ : if  $T^k$  is rank-1 then  $C(T^k) = C(T)$ . Hence if  $T$  is non-rigid then the rank of  $T^k$  is at least 2. (Contrast example (v).)

*Proof of (A).* If  $\mathcal{F}$  is a  $T$ -invariant sub-sigma-algebra then  $\mathcal{F}$  is invariant under any power of  $T$  and hence any weak limit of powers of  $T$ ; thus the group homomorphism  $\gamma$ , where  $\gamma(S) \triangleq S|_{\mathcal{F}}$ , is well defined.

The stated injectivity of  $\gamma$  is seen as follows: since  $F$  is ergodic on a non-atomic space,  $\text{Id} = F^s = \text{note } \gamma(T^s)$  implies  $s = 0$ . So the kernel of  $\gamma$  is trivial. Suppose  $T$  were a group extension of  $F$  by some group  $G$  acting by, say, left multiplication. By letting  $G$  act on the fibres over  $\mathcal{F}$  by right multiplication, we get a copy of  $G$  in  $C(T)$  and this copy is in the kernel of  $\gamma$ . So  $G$  is the trivial group and  $F = T$ .

*Proof of (B).* Every  $S \in C(T)$  is invertible so  $T$  cannot be a proper factor of itself and therefore  $\psi \circ \phi$  is an automorphism.

*Proof of (C).* Suppose we could find a transformation  $R: Y \rightarrow Y$  and an isomorphism  $X \cong Y \times Y$  such that  $T = R \times R$ . Any power of  $T$ , then, is a cartesian square and consequently so are weak limits of such. Were the coordinate-flip map  $(y, y') \mapsto (y', y)$  a cartesian square,  $Y$  would be a one point space.

That  $T$  cannot be a cartesian square implies that, if  $T$  has a factor of the form  $R_1 \times R_2$ , then  $R_1$  and  $R_2$  are relatively prime. For if  $R_1$  has a factor, call it  $R$ , which is isomorphic to a factor of  $R_2$ , then  $R_1 \times R_2$  has a factor of the form  $R \times R$ . Thus  $T$  has a factor  $R \times R$ . By a coding argument one can show that factors of rank-1 are themselves rank-1. Hence this  $R \times R$  is rank-1, contradicting the previous paragraph.

Now suppose  $C(T)$  is trivial and  $T$  is isomorphic to some non-trivial cartesian product  $R_1 \times R_2$ . Then  $\text{Id}_1 \times R_2 \in C(T)$  so there exists  $s_1 \neq 0$  such that  $\text{Id}_1 \times R_2 = T^{s_1}$ . Thus  $(R_1)^{s_1} = \text{Id}_1$ . Similarly, there exists  $s_2 \neq 0$  such that  $(R_2)^{s_2} = \text{Id}_2$ . Then  $s_1 s_2 \neq 0$  yet  $T^{s_1 s_2} = \text{Id}$ .

*Proof of (D).* Fix  $k \geq 2$ . For any transformation  $T$  we automatically have the three solid inclusions shown below:

$$\begin{array}{ccc} C(T) & \subset & C(T^k) \\ \cup & & \cap \\ \text{WCl}(T) & \supset & \text{WCl}(T^k) \end{array}$$

If  $T^k$  is rank-1, the dashed inclusion holds and therefore  $C(T^k) = C(T)$ . But  $T \in C(T^k)$  so  $C(T^k)$  is uncountable. Since  $C(T)$  is uncountable,  $T$  is rigid.  $\square$

*Remark.* The referee mentions that the first half of corollary (C) follows from known spectral results: A rank-1 must have simple spectrum (Baxter). A cartesian square has spectral multiplicity at least two. Hence a rank-1 cannot be a cartesian square.

More generally, no transformation of the form  $R^l \times R^k$  can be rank-1. If it were, then since  $R \times R \in C(R^l \times R^k)$  there is a sequence  $\{n_i\}$  such that

$$(R^l)^{n_i} \rightarrow R \quad \text{and} \quad (R^k)^{n_i} \rightarrow R.$$

Hence  $R^{ln_i k} \rightarrow R^k$  and  $R^{kn_i l} \rightarrow R^l$  so  $R^k = R^l$ . But a cartesian square cannot be rank-1.

*Examples.* Non-rigid rank-1 transformations are not well understood. The only types of examples known to the author are the following:

- (i) Ornstein's (or any) rank-1 mixing [9].
- (ii) Friedman and Ornstein's partial mixing transformation [2] but done in rank-1.
- (iii) Chacón's weak mixing transformation [6].
- (iv) The following skew product  $T$ . Let  $F: X \rightarrow X$  be a rotation of some discrete space  $X = \{0, 1, \dots, k - 1\}$ . Let  $S: Y \rightarrow Y$  be any non-rigid rank-1. Define  $T: X \times Y \rightarrow X \times Y$  by

$$T(x, y) \triangleq \begin{cases} (Fx, Sy) & \text{if } x = '0'; \\ (Fx, y) & \text{otherwise.} \end{cases}$$

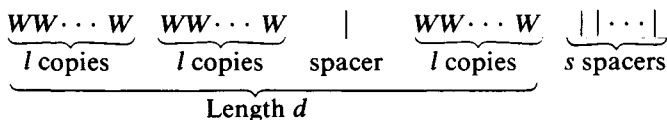
Transformations (i) and (iii) have minimal self-joinings, see [8] and [6], respectively. Transformation (iv) does not, possessing a discrete rotation as factor. Can more complex factors appear?

Turning our attention in the other direction, we give two examples of the kind of behaviour one can obtain by building-in rigidity.

(v) *A weak mixing transformation  $T$  whose non-zero powers are rank-1:* The weak mixing is the issue – otherwise an irrational rotation would serve. Define the function  $f: \mathbb{N} \rightarrow \mathbb{N}$  by  $f(n) \triangleq n! - 1$ . Then  $f$  has the property:

- (\*) For each  $n: \forall d \in \mathbb{Z} \exists s \in \{0, 1, \dots, f(n)\}$  such that  $d + s$  is relatively prime to each of  $2, 3, \dots, n$ .

We build  $T$  by cutting and stacking. At a fixed stage  $n$ , let  $W$  and  $W'$  denote the  $n$ -block and the  $(n + 1)$ -block, respectively, of heights  $h$  and  $h'$ . We build  $W'$  as a concatenation of copies of  $W$  and spacers.



Let  $d$  be the length  $3lh + 1$ . We pick  $l$  so that  $l > nh$  and also so that  $(f(n) + 1)/d < \frac{1}{2}$ . By (\*) we can pick an  $s \leq f(n)$  so that  $h' \triangleq 3lh + 1 + s$  is relatively prime to  $2, 3, \dots, n$ . The amount of mass we add as spacer at stage  $n$ , is  $(s + 1)/h'$ . Since this is less than  $\frac{1}{2}$ , a summable function of  $n$ , our cutting and stacking indeed does define a transformation on a probability space.  $T$  is seen to be weak-mixing by the standard argument showing Chacón's transformation weak-mixing.

Why is each  $T^k$  rank-1? We use the following criterion, equivalent to the definition given previously, to show a transformation is rank-1: Given any  $\epsilon$  and any partition  $Q$  there is a Rohlin stack such that  $Q$  is  $\epsilon$ -refined by the partition whose atoms are the stack's column levels and the stack complement.

Let  $B$  denote the base of the  $n$ -stack. From the preceding figure,  $\mu(B \sim T^h B) < 1/l < 1/nh$ . Let

$$A \triangleq B \cap (T^{-h})(B) \cap \dots \cap (T^{-h})^{k-1}(B).$$

Then  $\mu(B \sim A) \leq (k-1) \cdot \mu(B \sim T^h B)$  and we can have chosen  $n$  large enough that this latter is less than  $\varepsilon/2h$ . Also,  $T^{hj}(A) \subset B$  for any  $0 \leq j < k$ . Thus  $T^j(A)$  is contained in the  $B$ -stack  $\bigcup_{i=0}^{h-1} T^i B$ , for each  $0 \leq j < hk$ . We may assume  $n > k$  and so  $k$  is relatively prime to  $h$ . Consequently the sets  $A, T^k(A), \dots, (T^k)^{h-1}(A)$  are disjoint (since they are contained in different levels of the  $B$ -stack) and hence form a Rohlin stack for  $T^k$ . The partition distance from the column-level partition of this stack to that of the  $B$ -stack is less than  $2 \cdot h \cdot \mu(B \sim A) < 2h \cdot \varepsilon/2h = \varepsilon$ .

*Question.* A special case of a result in [8] is that for any rank-1 mixing  $T$ ,  $\text{rank}(T^k) = k$ . Is this true for Chacón’s transformation or, more generally, for any weak-mixing rank-1 transformation which is non-rigid? Under what circumstances (one can ask this for a general finite rank  $T$ ) must the function  $k \mapsto \text{rank}(T^k)$  be monotonic on those values  $k$  for which  $T^k$  is ergodic?

(vi) *A rank-1 weak mixing  $T$  which is a countable cartesian product  $R_1 \times R_2 \times \dots$ .* First let us address building a weak-mixing rank-1 product  $R_1 \times R_2$ . It is well known that one can build such a beast and A. del Junco remarks to this effect, in [4]. Having not seen a construction in print, we sketch one way to proceed. Pick a sequence  $\varepsilon_n \rightarrow 0$ . We build  $R_1$  and  $R_2$  simultaneously by cutting and stacking much as in example (v). For  $i = 1, 2$  let  $B_i(n)$  denote the base of the  $n$ -stack, of  $R_i$ , with height  $h_i(n)$ . We can arrange that for each  $n$  and  $i$ :

- (\*)  $h_1(n)$  and  $h_2(n)$  are relatively prime;
- (\*\*)  $\mu_i(B_i(n) \sim R_i^{h_i(n)}(B_i(n))) < \varepsilon_n/[p(n)]^2$ ;

where  $p(n) \triangleq h_1(n) \cdot h_2(n)$ . On the product space we can find, for each  $n$ , a set  $A \subset B_1 \times B_2$  satisfying

$$(0.3) \quad \mu_1 \times \mu_2([B_1 \times B_2] \sim A) < \varepsilon/p.$$

Here, the terms  $A, B_i, \varepsilon$ , and  $p$  denote  $A(n), B_i(n), \varepsilon_n$ , and  $p(n)$  respectively.

We make  $A$  as follows. Let  $A = A_1 \times A_2$  where

$$A_1 \triangleq B_1 \cap (R_1^{-h_1})(B_1) \cap \dots \cap (R_1^{-h_1})^{h_2}(B_1).$$

Thus, by (\*\*),  $\mu_1(B_1 \sim A_1) < h_2 \cdot \varepsilon/p^2 < (\varepsilon/2)/p$ . Now define  $A_2$  by reversing the roles of ‘1’ and ‘2’. Hence  $A \subset B_1 \times B_2$  and (0.3) holds.

For any  $k, 0 \leq k < p$ , notice that  $(R_1 \times R_2)^k(A) \subset R_1^{k_1}(B_1) \times R_2^{k_2}(B_2)$  where  $k_i \triangleq k \pmod{h_i}$ . So, by (\*), the first  $p$  iterates of  $A$  are disjoint. Moreover, these  $p$  iterates fill up most of the first  $p$  iterates of  $B_1 \times B_2$ .

But, letting  $n$  range over  $\mathbb{N}$ , the first  $p(n)$  iterates of  $B_1(n) \times B_2(n)$  generate the product space’s sigma-algebra. Consequently, so do the first  $p(n)$  iterates of  $A(n)$ . Hence  $R_1 \times R_2$  is indeed rank-1.

It is worth remarking that we built rigidity into the  $R_i$ , and the weak-closure theorem, corollary (C), comforts us in that we had no choice. But it also implies something which is less apparent from the construction; that  $R_1 \neq R_2$ .

In general, construct  $R_1, R_2, R_3, \dots$  simultaneously so that at each stage  $n$ :  $h_1(n), \dots, h_n(n)$  are relatively prime and  $(**)$  holds for  $i=1, \dots, n$  where  $p(n) \triangleq \prod_{i=1}^n h_i(n)$ . Thus each  $R_1 \times \dots \times R_n$  is rank-1 weak-mixing.

The countable product  $R_1 \times R_2 \times \dots$  is weak mixing and seen to be rank-1 as follows. Suppose  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$  is a sequence of factor algebras of a transformation  $T$  such that  $\bigvee_1^\infty \mathcal{F}_n$  is the whole sigma-algebra. If  $T|_{\mathcal{F}_n}$  is rank-1 for each  $n$ , then the ‘ $\epsilon$ -refining’ definition of rank-1, given in example (v), shows that  $T$  is rank-1: Any partition  $Q$  can be approximated as well as desired by some  $Q'$  living in some  $\mathcal{F}_n$ .

*Remark.* By corollary (C) of the weak-closure theorem,  $R_1 \times R_3 \times R_5 \times \dots$  must be relatively prime to  $R_2 \times R_4 \times \dots$ . The  $\{R_i\}_1^\infty$  have the curious property that for any set  $I$  of indices, the transformations

$$\prod_{i \in I} R_i \quad \text{and} \quad \prod_{j \in \mathbb{N} \setminus I} R_j$$

are relatively prime.

1. To prepare for the proof of the weak closure theorem, we establish a few lemmas.

(1.1) *Definition.* Let  $T: X \rightarrow X$  be ‘cutting and stacking’ rank-1 with generating partition  $P$ . Denote the  $n$ -block by  $W_{(n)}$  and its height by  $h_n$ . Thus  $W = W|_0^h$  is a  $P$ - $h$ -word where, as will be usual, we have suppressed the subscript.

Given  $S \in C(T)$  we can approximate  $S$  by a finite code  $\mathcal{C}$ . That is, given  $\epsilon$  we can find  $l$  such that

$$(1.1.1) \quad \bigvee_{i=-l}^l T^i P \supseteq_\epsilon S^{-1}(P).$$

Our code is a map from  $P[-l, l]$ -words to the alphabet  $P$ . Let  $\text{CodeLen}(\mathcal{C})$  mean the quantity  $2l+1$ . (1.1.1) and the ergodic theorem say that for a.e.  $x \in X$ ,  $\bar{d}(\mathcal{C}(x), Sx) \leq \epsilon$ . Define  $\text{ErrorRate}(\mathcal{C}) \triangleq 2\epsilon$ . We say that a substring  $x|_j^{j+b}$  codes well (under  $\mathcal{C}$ ) if

$$\bar{d}(\mathcal{C}(x|_j^{j+b}), (Sx)|_j^{j+b}) < \text{ErrorRate}(\mathcal{C}).$$

Following the usual convention, the writing of  $\mathcal{C}(x|_j^{j+b})$  shall tacitly imply that  $b \gg \text{CodeLen}(\mathcal{C})$ . Consequently, we can view  $\mathcal{C}(x|_j^{j+b})$  as a string of length  $b$  by harmlessly absorbing the end effects of the code into its error rate.

Because we gave ourselves a bit of room by defining the error rate to be twice  $\epsilon$ , an application of the ever-popular ergodic theorem yields the:

(1.2) **STANDARD CODING LEMMA.** Fix some code  $\mathcal{C}$ . For a set of  $x \in X$  of full measure,  $\forall \delta', \forall$  large  $h$ : Cover, in any way, at least  $(1 - \delta')$ -percent of  $x|_0^\infty$  by disjoint substrings of length  $h$ . Then at least  $(1 - 2\delta')$ -percent of  $x|_0^\infty$  is covered by those, of the above substrings, which code well under  $\mathcal{C}$ .

*Proof.* Standard.

Now let  $\mathbf{C} = \{\mathcal{C}^j\}_{j=1}^\infty$  be a collection of codes such that  $\text{ErrorRate}(\mathcal{C}^j) \rightarrow_j 0$ . Let  $\mathcal{A}$  be the algebra of  $T$ ,  $P$ -cylinder sets. The algebra  $\mathcal{A} \vee S^{-1}(\mathcal{A})$  is countable and so, for a set of full measure of  $x \in X$ ,  $x|_0^\infty$  hits each set in  $\mathcal{A} \vee S^{-1}(\mathcal{A})$  with the appropriate

frequency. We may also assume that for these same  $x$ , the standard coding lemma works on  $x$  for each code in the countable collection  $\mathbf{C}$ . Consequently, we can pick a point  $\check{x}$ , then an  $x \in S^{-1}(\{\check{x}\})$ , such that both lie in this set of full measure.

(\*) For the remainder of this section we consider  $T, P, S, x$  and  $\check{x}$  as fixed. The word *code* henceforth means some code in  $\mathbf{C}$ . Also, when considering doubly infinite  $T, P$ -names such as  $y, z \in X$ , agree to interpret  $\bar{d}(y, z)$  as  $\bar{d}(y|_0^\infty, z|_0^\infty)$ .

(1.3) LEMMA. For any sequence of integers  $s_j$ :

$$(1.3.1) \quad T^{s_j} \xrightarrow[j]{\text{weakly}} S \Leftrightarrow \bar{d}(T^{s_j}(x), \check{x}) \xrightarrow[j]{} 0.$$

*Proof.* The left hand side of the above is equivalent to:

$$\text{For each cylinder set } A: \mu(T^{-s_j}(A) \Delta S^{-1}(A)) \rightarrow 0.$$

For any such  $A$  and any  $j$ , the set  $T^{-s_j}A \Delta S^{-1}A$  is a member of our algebra  $\mathcal{A} \vee S^{-1}(\mathcal{A})$ . To prove the ( $\Leftarrow$ ) direction, pick  $l$  such that  $A$  is an atom of  $\bigvee_{-l}^{l-1} T^i P$  and let  $A|_{-l}^l$  denote the  $P$ -word corresponding to  $A$ .

$$\begin{aligned} \mu(T^{-s_j}A \Delta S^{-1}A) &= \text{freq} \{i \in \mathbb{N}: T^i x \in T^{-s_j}A \text{ XOR } T^i x \in S^{-1}A\} \\ &= \text{freq} \{i \in \mathbb{N}: T^{s_j}(x)|_{i-l}^{i+l} = A|_{-l}^l \text{ XOR } \check{x}|_{i-l}^{i+l} = A|_{-l}^l\} \\ &\leq \text{freq} \{i \in \mathbb{N}: T^{s_j}(x)|_{i-l}^{i+l} \neq \check{x}|_{i-l}^{i+l}\} \\ &\leq 2l \cdot \bar{d}(T^{s_j}(x), \check{x}). \end{aligned}$$

Now send  $j \rightarrow \infty$ . The proof of the ( $\Rightarrow$ ) direction is similar. □

Recall that we let  $W_{(n)}$  or just  $W$  denote the word which is the  $n$ -block. Its length is  $h_n$  or just  $h$ . For an integer  $s$  measuring the length of something connected with  $W$ , we let  $s\%$  denote the number  $s/h$ .

For any  $y \in X$  we say that a substring  $y|_i^{i+h}$  is a  $W$ -copy if it is an  $n$ -block i.e. if  $T^i y$  is a point in the base of the  $n$ -stack.

For the sake of contradiction, assume henceforth that  $S \notin \text{WCl}(T)$ .

(1.4) SHIFT LEMMA.  $\exists$  constants  $e', \sigma > 0$  such that for any code  $\mathcal{C}$  with Error-Rate( $\mathcal{C}$ )  $< e', \forall$  large  $n$ , whenever a shift value  $s \in |_{-h}^h$  is such that:

(in the case  $s \geq 0$ ):

$$\bar{d}(\mathcal{C}(W|_s^h), W|_0^{h-s}) < e',$$

(in the case  $s \leq 0$ ):

$$\bar{d}(\mathcal{C}(W|_0^{h+s}), W|_{-s}^h) < e',$$

then  $|s\%| > 2\sigma$ .

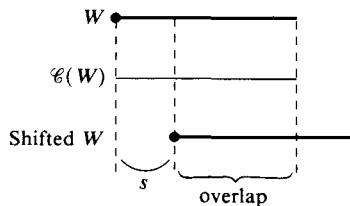


FIGURE 1. If the  $\bar{d}$  distance between  $\mathcal{C}(W)$  and the shifted  $W$ -copy is small on their overlap, then the shift,  $s$ , measured as a fraction of the length of the  $n$ -block, must be larger than a certain constant,  $2\sigma$ . Here  $s$  is drawn positive.



*Proof.* Since  $S \notin \text{WCl}(T)$  there exists  $\varepsilon$  for which

$$4\varepsilon < \bar{d}(T^{-s}x, \check{x}), \quad \forall s \in \mathbb{Z}.$$

If the lemma is false, then we can choose a code  $\mathcal{C}$ ,  $\text{ErrorRate}(\mathcal{C}) \stackrel{\cong}{=} e < \varepsilon$ , and then a large  $n$  and non-negative (without loss of generality) shift  $s \in \mathbb{Z}$  such that  $s \% h < \varepsilon$  and

$$\bar{d}(\mathcal{C}(W|_s^h), W|_0^{h-s}) < e < \varepsilon.$$

We may assume  $n$  sufficiently large that  $\delta < \varepsilon$  where  $1 - \delta \stackrel{\cong}{=} \mu(n\text{-stack})$ . Suppose  $x|_i^{i+h}$  is a  $W$ -copy. Then

$$\begin{aligned} & \bar{d}((T^{-s}x)|_i^{i+h}, \mathcal{C}(x|_i^{i+h})) \\ & \leq s \% \cdot (\bar{d} \text{ error on } |_i^{i+s}) + [1 - s \%](\bar{d} \text{ error on } |_{i+s}^{i+h}) \\ & \leq \varepsilon \cdot 1 + 1 \cdot e < 2\varepsilon. \end{aligned}$$

Since  $(1 - \delta)$ -percent of  $x|_0^\infty$  is covered by  $W$ -copies,

$$\bar{d}(T^{-s}x, \mathcal{C}x) \leq \delta \cdot 1 + (1 - \delta) \cdot 2\varepsilon < 3\varepsilon.$$

So  $\bar{d}(\mathcal{C}x, \check{x}) < e < \varepsilon$  produces  $\bar{d}(T^{-s}x, \check{x}) < 4\varepsilon$  and a contradiction. □

Agree to regard  $\sigma$  as a primordial constant, ‘known in advance’. In the sequel, *code* means a code with error rate less than  $e'$ . We now can forget the existence of  $e'$  and restate the lemma informally as:

*For any code and for all large  $n$ : If the  $n$ -block  $W$  codes well on the overlap to a shift of itself, then this shift must exceed an a priori percentage of  $W$ .*

(1.5) *Definition.* For a fixed implicit  $n$ ,  $W = W|_0^h$  represents the  $n$ -block. For any  $i, p \in \mathbb{Z}$  let  $i \oplus p$  mean the number  $i + p \pmod{h}$ . Let  $W \oplus p$  mean the  $h$ -word defined by

$$(W \oplus p)|_i \stackrel{\cong}{=} W|_{i \oplus p} \quad \text{for each } i \in \mathbb{Z}.$$

For any  $\varepsilon$  we say that  $W$  is  $\varepsilon$ -periodic (of period  $p$ ) if  $\bar{d}(W, W \oplus p) < \varepsilon$ . Saying  $W$  has periodicity  $p$  will mean  $W$  is  $\varepsilon$ -periodic for some as yet unspecified  $\varepsilon$ .

Let  $N$  denote the integer constant for which  $N + 1 \geq 1/\sigma > N$ .

(1.6) **PERIODICITY EXTENSION LEMMA.** *Given  $\varepsilon$ , there exists  $\varepsilon'$  and codes  $\{\mathcal{C}_i\}_1^N$  such that  $\forall$  large  $h$ : Suppose we have an  $h$ -word  $W$ , a  $p \in \mathbb{Z}$ , and an  $r \in \mathbb{N}$  with  $r \% h \geq \sigma$ . Suppose that when  $r' \stackrel{\cong}{=} r$  (or  $r' \stackrel{\cong}{=} -r$ ) these two conditions hold:*

- (i)  $\bar{d}(W|_0^r, (W \oplus p)|_0^r) < \varepsilon'$ ;
- (ii) for each  $i$ ,  $\bar{d}(\mathcal{C}_i(W), W \oplus r') < \text{ErrorRate}(\mathcal{C}_i)$ .

Then  $\bar{d}(W, W \oplus p) < \varepsilon/2$ .

*Remark.* Informally, this says that periodicity on a sufficiently long piece of  $W$  can be extended to all of  $W$ . Although the error rate will get worse, we will not lose control of it.

*Proof.* There is no loss of generality in assuming  $r' = r$ . Setting  $\varepsilon_0 \stackrel{\cong}{=} \varepsilon/2$  we inductively pick the codes as  $i = 1, 2, \dots, N$ .

At stage  $i$ : Pick a positive  $e_i < \varepsilon_{i-1}/2$ . Then choose a code  $\mathcal{C}_i$  with error rate less than  $\sigma e_i$ . Set  $k_i \stackrel{\cong}{=} \text{CodeLen}(\mathcal{C}_i)$ . Finally, pick an  $\varepsilon_i$  so small that  $k_i \varepsilon_i + 2e_i < \varepsilon_{i-1}$ .

The induction done, set  $\varepsilon' \triangleq \varepsilon_N$ . Suppose, now, we are given an  $h$  sufficiently enormous that  $\sigma h \gg \max_i k_i$  and, consequently, we may absorb the ‘end effects’ of our codes into their rates. First, we copy the periodicity on  $|'_0$  to  $|'_r$  using code  $\mathcal{C}_N$ :

$$\bar{d}(\mathcal{C}_N(W|'_0), \mathcal{C}_N(W \oplus p|'_0)) \leq k_N \cdot \bar{d}(W|'_0, W \oplus p|'_0) < k_N \varepsilon_N.$$

From (ii) we get  $\bar{d}(W|^{2r}_r, \mathcal{C}_N(W|'_0)) < \varepsilon_N$  and  $\bar{d}(\mathcal{C}_N(W \oplus p|'_0), W \oplus p|^{2r}_r) < \varepsilon_N$ . By the triangle inequality

$$\bar{d}(W|^{2r}_r, W \oplus p|^{2r}_r) < \varepsilon_N + k_N \varepsilon_N + \varepsilon_N < \varepsilon_{N-1}.$$

We can inductively continue to show for each  $i = 1, \dots, N$  that

$$\bar{d}(W|^{(i+1)r}_{ir}, W \oplus p|^{(i+1)r}_{ir}) < \varepsilon_{N-i} \leq \varepsilon_0.$$

This says that  $\bar{d}(W, W \oplus p) < \varepsilon_0$  – or rather, it does if  $r$  divides  $h$ . To handle a general  $r$ , we choose the  $\varepsilon_i$  inductively as before but with  $\varepsilon_0$  replaced by  $(\varepsilon_0/2)^2$ . This yields, for each  $i$

$$(1.6.1) \quad \bar{d}(W|^{(i+1)r}_{ir}, W \oplus p|^{(i+1)r}_{ir}) < (\varepsilon_0/2)^2.$$

Let  $M$  be the integer  $Mr < h < (M+1)r$ . Either  $(h - Mr)\%$  exceeds  $\varepsilon_0/2$  or not. In either case (by Fubini, in the former)

$$(1.6.2) \quad [h - Mr] \cdot \bar{d}(W|^{h}_{Mr}, W \oplus p|^{h}_{Mr}) < h \cdot \frac{\varepsilon_0}{2}.$$

Since  $(\varepsilon_0/2)^2$  is less than  $\varepsilon_0/2$ , (1.6.1) implies

$$Mr \cdot \bar{d}(W|^{Mr}_0, W \oplus p|^{Mr}_0) < Mr \cdot \frac{\varepsilon_0}{2} < h \frac{\varepsilon_0}{2}.$$

This together with (1.6.2) yields the desired  $\bar{d}(W, W \oplus p) < \varepsilon_0 = \varepsilon/2$ . □

(1.7) *Proof of weak-closure theorem.* Using an  $\varepsilon$  to be specified later, the periodicity extension lemma hands us a collection of  $N$  codes and the number  $\varepsilon'$ . Let  $\mathcal{C}$  be an even better code having error rate  $e$  where  $2e/\sigma < \varepsilon'$ .

Pick positive  $\delta$  and  $\gamma$  to be specified later. Lastly, choose an  $n$  so that  $h \triangleq h_n$  is sufficiently large to work for the periodicity extension lemma. Suppose that  $x|^{i+h}_i$  and  $x|^{i+h+g+h}_{i+h+g}$  are two successive  $W$ -copies on  $x$ . Suppose there is a  $W$ -copy  $x|^{j+h}_j$  ‘lying below’ i.e. with  $i \leq j < i + h + g$ . Say that  $W$ -copy  $x|^{i+h}_i$  is *good* if:

(1.7.1)  $g\% < \gamma$  and each of the words  $x|^{i+h}_i$ ,  $x|^{i+h+g+h}_{i+h+g}$ , and  $x|^{j+h}_j$  code well under our  $N+1$  codes.

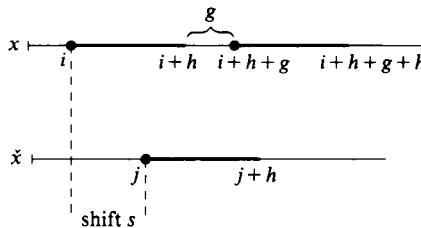


FIGURE 2. Picture of a good  $W$ -copy  $x|^{i+h}_i$ . We call  $s \triangleq j - i$  the *shift* associated with  $x|^{i+h}_i$  and  $g$  the associated *gap*. We call  $x|^{i+h+g+h}_{i+h+g}$  its *successor*  $W$ -copy (which may or may not be good). (A  $W$ -copy is shown as a thick bar with a large dot indicating its first coordinate.)

One may have chosen  $n$  sufficiently large that:

(1.7.2) At least  $(1 - \delta)$ -percent of  $x|_0^\infty$  is covered by good  $W$ -copies.

For we can have chosen  $n$  large enough that  $(1 - \delta')$ -percent of  $x|_0^\infty$  and  $\check{x}|_0^\infty$  are covered by  $W$ -copies, where  $\delta'$  is some number  $\delta' \ll \delta$  and  $\delta' \ll \gamma^2$ . Thus most  $W$ -copies on  $x|_0^\infty$  have a  $W$ -copy 'lying below' on  $\check{x}$  and, by a Fubini argument, have a gap less than  $\gamma$ . Moreover, we can have insisted  $h (=h_n)$  be so large that the standard coding lemma works (for our chosen  $\delta$ ) for all  $N + 1$  codes simultaneously. Thus (1.7.2) holds.

Since the good  $x|_i^{i+h}$  of figure 2 codes well, the shift lemma says that  $s\% > 2\sigma$ . Because its successor codes well,  $1 - s\% + g\% > 2\sigma$ . We may assume  $\gamma < \sigma$ . Therefore  $\sigma < s\% < 1 - \sigma$ .

*Periodicity suffices.* We make a simplifying assumption to be removed later: Suppose there is a good  $W$ -copy with gap zero. Pick one such, henceforth called the *reference* ( $W$ -copy), and denote its associated shift by  $r$ . That the reference is good tells us that  $\bar{d}(\mathcal{C}(W), W \oplus (-r)) < \epsilon$  (this is shown pictorially in figures 4 and 5). Our theorem is proved, should one be able to insure that:

(1.7.3) For each  $g$ , the gap associated to a good  $W$ -copy: The  $n$ -block,  $W$ , has  $\epsilon$ -periodicity of period  $g$ .

*Proof of sufficiency of (1.7.3).* Let  $0 \leq i_1 < i_2 < \dots$  be the first coordinates of the successive  $W$ -copies, good or not, on  $x|_0^\infty$ . Let  $g_l$  denote the gap between  $x|_{i_l}^{i_l+h}$  and its successor. If  $x|_{i_l}^{i_l+h}$  is good we say, by abuse of notation, that  $l$  is good. We

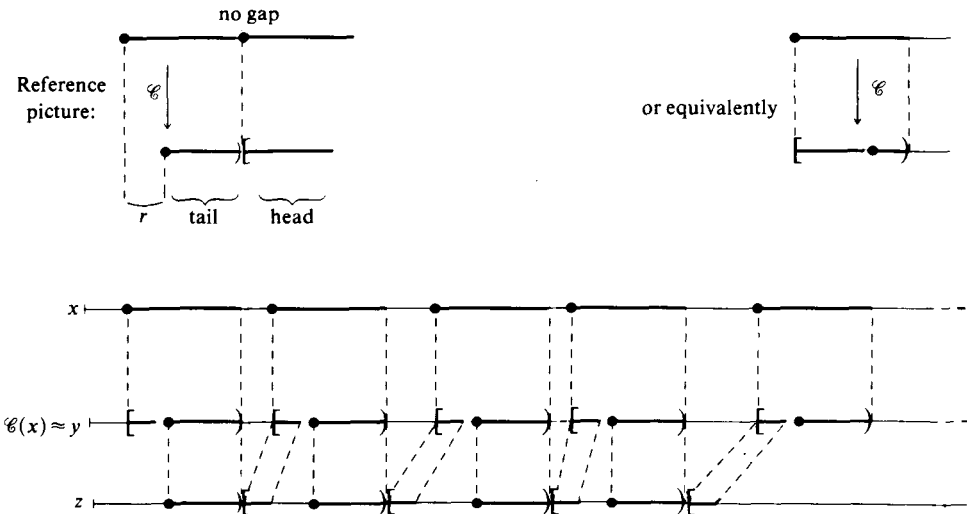


FIGURE 3. The reference says that  $W$  codes well to a rotation of itself. Thus we can split  $W$ , into a *tail*  $W|_0^{h-r}$ , and a *head*,  $W|_{h-r}^h$ . Restated,  $W$  codes well to itself with the head and tail transposed. We construct a name  $y$ ,  $\bar{d}$ -close to  $\mathcal{C}(X)$ , simply by taking  $x$  and transposing the head, and tail of each  $W$ -copy on  $x$ .

To make  $z$ , take  $y$  and pull each head back over the gap behind it and glue it to the preceding tail. For most heads this shift in position makes only a small  $\bar{d}$ -change; for most gap lengths, the word  $W$  - hence the head of  $W$ , is  $\bar{d}$ -periodic of that length.

synthesize a half-infinite  $P$ -name  $y = y|_0^\infty$ . For  $l = 1, 2, \dots$  define

$$y|_i^{i+h} \triangleq W \oplus -r \stackrel{\text{note}}{=} W|_{h-r}^h, W|_0^{h-r}.$$

Set  $y$  equal to  $\mathcal{C}x$  elsewhere i.e.  $y|_{i+h}^{i+h+g_l} \triangleq \mathcal{C}(x)|_{i+h}^{i+h+g_l}$ . Then

$$\bar{d}(y, \mathcal{C}(x)) < \bar{d}(W \oplus -r, \mathcal{C}(W)) < e.$$

So

$$\bar{d}(y, \check{x}) \leq \bar{d}(y, \mathcal{C}x) + \bar{d}(\mathcal{C}x, \check{x}) \leq e + e < 2e.$$

Fix some  $l$  and let  $i, g, I$  denote  $i_i, g_i, i_i + h$  respectively. Recall that  $y|_{I+g}^{I+g+r} = W|_{h-r}^h$ . If our chosen  $l$  is good then (1.7.3) gives  $\bar{d}(W, W \oplus g) < \varepsilon$  and consequently

$$\bar{d}(W|_{h-r}^h, y|_I^{I+r}) < (\varepsilon + g\%) / r\% < (\varepsilon + \gamma) / \sigma \stackrel{\text{WLOG}}{<} 2\varepsilon / \sigma.$$

We construct a  $P$ -name  $z = z|_0^\infty$ . For each good  $l$ , define  $z|_{i+h}^{i+h+r} \triangleq W|_{h-r}^h$ . Elsewhere, let  $z$  equal  $y$ . Then

$$\bar{d}(z, y) < (2\varepsilon / \sigma) \cdot r\% < 2\varepsilon / \sigma.$$

But a glance at figure 3 shows that for good  $l$

$$z|_{i+r}^{i+r+h} = W = x|_i^{i+h}.$$

Thus  $\bar{d}(T^{-r}x, z) < \delta$ , by (1.7.2). Consequently

$$\bar{d}(T^{-r}x, \check{x}) < \delta + (2\varepsilon / \sigma) + 2\varepsilon \stackrel{\text{WLOG}}{<} 5\varepsilon / \sigma.$$

But we could have initially chosen  $\varepsilon$  sufficiently small that

$$5\varepsilon / \sigma < \inf_{s \in \mathbb{Z}} \bar{d}(T^{-s}x, \check{x}).$$

This gives a contradiction, and hence (1.7.3) suffices.

*Obtaining partial periodicity.* For the remainder of this discussion, fix some good  $W$ -copy, henceforth called the *object*  $W$ -copy, and let  $g$  and  $s$  denote its associated gap and shift, respectively. We will obtain the desired periodicity  $\bar{d}(W, W \oplus g) < \varepsilon$  by comparing the constraints placed on  $W$  by the object and reference  $W$ -copies. Let  $k$  denote the odd number  $\text{CodeLen}(\mathcal{C})$ . We now wrap both pictures into circles. For the reference picture, we simply identify the two  $W$ -copies in the upper line.

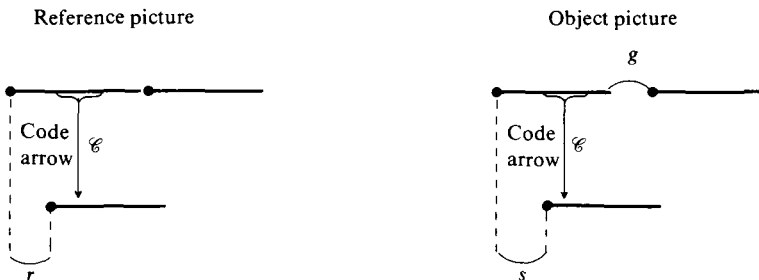


FIGURE 4. In both pictures the upper line codes well under  $\mathcal{C}$  to the lower line. For each position in the lower line (of either picture), the code ‘guesses’ what letter is there by looking at the  $k$ -word above centred over that position. The ‘guess’ disagrees with what is actually in the lower line, with frequency less than  $e$ .

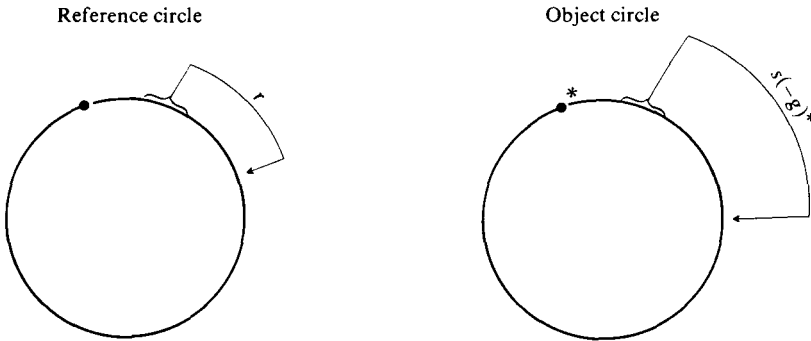


FIGURE 5. The reference and object pictures are each wrapped into circles of circumference  $h$ . The reference circle has caliper length  $r$ . The object circle has a large asterisk replacing the gap. Its caliper has varying length;  $s - g$  when the caliper straddles  $*$  and length  $s$  when it does not straddle  $*$ . As drawn here, the caliper is not straddling the  $*$ , that is, its code end and its arrow end are not on opposite sides of the  $*$ . So, as drawn here, the caliper length is  $s$ .

Since the lower line is also a  $W$ -copy,  $\mathcal{C}$  now maps the circle to itself and the code arrow gets bent into a sort of code ‘caliper’ which maps a  $k$ -word to a letter  $r$  positioned clockwise around the circle. The caliper has a ‘code end’, of length  $k$ , and an ‘arrow end’, of length 1. If we slide the code end once around the circle, one position at a time, then the letter the arrow end writes down agrees with the letter the arrow is pointing to, at least  $(1 - e)$ -percent of the time i.e.  $\bar{d}(\mathcal{C}(W), W \oplus -r) < e$ .

We wrap the object picture into a circle by identifying the two  $W$ -copies in the upper line of figure 4, discarding the gap. Without the presence of the gap, the caliper has length  $s - g$  or  $s$ , depending on whether or not it straddles the spot where the gap used to be. If we slide this caliper around its circle, we see errors at the arrow end with frequency less than  $e$ . (Actually, there are additional errors due to ‘end effects’ as the code end of the caliper slides over the asterisk. The frequency of these errors is of the order of  $2k\% + 2g\% < 2k\% + 2\gamma$ . Both  $k\%$  and  $\gamma$  can have been made arbitrarily small by having picked  $n$ , hence  $h$ , sufficiently large. Consequently, we can harmlessly absorb these end effects into the code error rate.)

Agree to refer to the calipers of the reference and object pictures as the  $r$ -caliper and  $s$ -caliper, respectively. We now superimpose the two circles. The calipers are

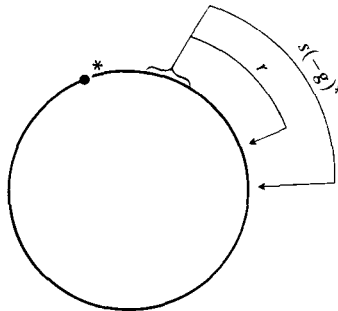


FIGURE 6. The calipers of the previous figure have been superimposed. Therefore, the two letters in  $W$ , pointed at by the two arrowheads, can fail to agree with frequency at most  $2e$ .

ganged at the code end. We slide this once around the circle. Since the arrows usually point to equal letters – with frequency better than  $1 - 2\epsilon$ , we see that the circle,  $W$ , has two complementary intervals of periodicity. These two intervals are, respectively, the set of positions of the  $s$ -arrow when the  $s$ -caliper straddles/does-not-straddle, the asterisk. So their percentage lengths are  $s\%$  and  $1 - s\%$ . Their periods are (measuring counterclockwise)  $p$  and  $p - g$ , respectively, where  $p \triangleq -(s - g - r)$ . The quantity  $p$  may be positive, negative, or zero.

*Extending, then comparing, periodicity.* What will be important is the length of the two complementary intervals of periodicity, not their locations on the circle. For notational convenience, let  $W|_0^s$  denote the interval with periodicity  $p$  and  $W|_s^h$  the interval with periodicity  $p - g$ .

Recall that, in figure 6, the letters in the positions indicated by the two arrows disagree with frequency less than  $2\epsilon$ , as we slide the ganged calipers once around the circle. Suppose we know that  $r\% \leq s\%$ . Then

$$\bar{d}(W|_0^r, W \oplus p|_0^r) < 2\epsilon/r\% < 2\epsilon/\sigma < \epsilon'.$$

Since the reference  $W$ -copy is good, the  $N$  codes we chose using the periodicity extension lemma each work well for the reference picture of figure 5. Hence we may appeal to this lemma, with  $r' \triangleq -r$ , and conclude that  $\bar{d}(W, W \oplus p) < \epsilon/2$ . Similarly, if  $r\% \leq 1 - s\%$ , we could pick a subinterval of  $W|_{h-s}^h$  of length  $r$  and extend its periodicity to obtain

$$\epsilon/2 > \bar{d}(W \oplus (p - g), W) \stackrel{\text{note}}{=} \bar{d}(W \oplus p, W \oplus g).$$

Beating one period against the other yields the desired

$$\bar{d}(W, W \oplus g) \leq \bar{d}(W, W \oplus p) + \bar{d}(W \oplus p, W \oplus g) < \epsilon.$$

*Final details.* The above argument lives if  $r\% \leq s\%$  and  $r\% \leq 1 - s\%$ . Or, if  $1 - r\% \leq s\%$  and  $1 - r\% \leq 1 - s\%$ , the argument persists by using  $r' \triangleq r$  in the periodicity extension lemma. In order to choose the reference copy so that one of these cases always holds, let  $s_1, s_2, \dots$  denote the shifts, à la figure 2, associated with the good  $W$ -copies along  $x|_0^\infty$ . Let  $l$  be an index such that

$$\min(s_l\%, 1 - s_l\%) = \inf_{1 \leq L < \infty} (\min(s_L\%, 1 - s_L\%)).$$

We choose this good  $W$ -copy,  $x|_{i_l}^{i_l+h}$ , to be our reference and set  $r \triangleq s_l$ .

But the associated gap,  $g_l$ , need not be zero. However, we can do all our previous arguments with  $W$  and  $h$  replaced by  $W'$  and  $h'$  where  $h' \triangleq h + g_l$  and  $W' \triangleq x|_{i_l}^{i_l+h'}$ . We can still draw the reference and object pictures of figure 4 – but now the ‘gap length’,  $g$ , of the object picture may be negative. Nonetheless, we still have  $|g\%| < \gamma$  and so we can do the arguments of *Obtaining partial periodicity* and *Extending, then comparing, periodicity* as before. The argument in *Periodicity suffices* also persists, modulo minor notational changes (in figure 3, for example, ‘heads’ of  $W$ -copies will be pushed forward over negative gaps, in going from  $y$  to  $z$ ). This completes the proof of the weak-closure theorem. □

*Remark.* This theorem is not true, in general, for finite rank mixing transformations; for instance, if  $S$  is rank-1 mixing and we set  $T \triangleq S^k$  for some  $k \geq 2$ . Since  $T$  is

mixing,  $\text{WCl}(T) = \{T^n\}_{-\infty}^{\infty}$ . Thus  $S \in C(T) \sim \text{WCl}(T)$ . One can easily show (e.g. [8]) that  $T$  is rank- $k$  and  $C(T) = \{S^n\}_{-\infty}^{\infty}$ .

2. In this section we show that any proper factor of a rank-1 transformation is rigid; hence, each non-atomic factor has uncountable commutant. (I am indebted to Dan Rudolph for several conversations on this result. Chris Bose, a doctoral student of M. Akcoglu at the University of Toronto, was able to prove this same result, using the techniques of our § 1.)

We start with rank-1 transformation  $R: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega, \mathcal{A}, \mu)$  with ‘cutting and stacking’ generating partition  $Q$ . Suppose we are given a partition  $P$  on  $\Omega$ . This determines the factor sigma-algebra  $\mathcal{F} \triangleq \bigvee_{-\infty}^{\infty} R^i P$ . Let  $T$  denote the factor transformation  $R|_{\mathcal{F}}$  and let  $X$  denote the factor space. We can view  $P, \mathcal{F}$  and  $\mu$  as living on  $X$  and so  $T: (X, \mathcal{F}, \mu) \rightarrow (X, \mathcal{F}, \mu)$ . We get the commutative diagram

$$\begin{array}{ccc} \Omega & \xrightarrow{R} & \Omega \\ \varphi \downarrow & & \downarrow \varphi \\ X & \xrightarrow{T} & X \end{array}$$

where  $\varphi$  represents the factor homomorphism.

To show  $T$  rigid it is enough to show:

(2.1)  $\exists$  a never-zero sequence  $\{s_i\}_1^{\infty}$  such that  $T^{s_i} \rightarrow \text{Id}$ .

*Proof* (sketch). The goal of this sketch is to define terms  $x, \check{x}, \mathcal{C}$ , and *good  $W$ -copy* which, in our factor context, will play analogous roles to the terms of the same name in § 1. We then will be able to appeal to the machinery of (1.7).

Without loss of generality, every  $z \in X$  is typical for the  $T, P$ -process. So, emulating lemma 1.3, we can choose some  $T, P$ -name  $z \in X$  and reduce the problem to showing: Given  $\varepsilon, \exists s \neq 0$  such that  $\bar{d}(T^s z, z) < 100\varepsilon$ . It suffices, then, to show:

(2.1')  $\forall \varepsilon, \exists P$ -names  $x$  and  $\check{x}$  and an  $s \neq 0$  such that  $\bar{d}(z, x) < \varepsilon$  and  $\bar{d}(z, \check{x}) < \varepsilon$  and  $\bar{d}(T^s x, \check{x}) < \varepsilon$ .

So, for the sake of contradiction, we assume there is some positive number  $\sigma'$  such that: For any sufficiently small  $\varepsilon$  and any  $P$ -names  $x$  and  $\check{x}$  such that  $\bar{d}(z, x) < \varepsilon$  and  $\bar{d}(z, \check{x}) < \varepsilon$ , then

(2.2)  $\inf_{s \in \mathbb{Z} - \{0\}} \bar{d}(T^s x, \check{x}) > \sigma'$ .

*Choosing  $\omega$  and  $\check{\omega}$ .* For the  $R, Q$ -process, let  $W'_n$  and  $h'_n$  denote the  $n$ -block and its length, respectively. Usually we suppress the subscript and write  $W'$  and  $h'$ . For any  $\omega \in \Omega$ , say that  $\omega|_i^{i+h'}$  is a  $W'$ -copy if  $R^i(\omega)$  is in the base of the  $n$ -stack.

It is convenient to defer the proof of the following fact: We can choose and fix two points  $\omega, \check{\omega} \in \Omega$  for which  $\varphi(\omega) = \varphi(\check{\omega})$  and the following holds.

(2.3)  $\forall \delta, \forall$  large  $n$ : At most  $\delta$ -percent of  $\omega|_0^{\infty}$  is covered by  $W'$ -copies  $\omega|_i^{i+h'}$  such that  $\check{\omega}|_i^{i+h'}$  is also a  $W'$ -copy.

In other words, rarely do  $W'$ -copies along  $\omega$  line up perfectly with  $W'$ -copies along  $\check{\omega}$ . We will prove this fact in §3 using the idea of self-joinings.

Making the names  $x$  and  $\check{x}$  and the code  $\mathcal{C}$ . Henceforth the points  $\omega$  and  $\check{\omega}$ , in  $\Omega$ , and the point  $z \triangleq \varphi(\omega) = \varphi(\check{\omega})$ , in  $X$ , are viewed as fixed. The factor homomorphism  $\varphi$  is a map from  $R, Q$ -names to  $T, P$ -names. It can be approximated arbitrarily well by finite codes. We will call such a code, a  $\varphi$ -code. There are but countably many  $\varphi$ -codes so we may assume that each  $\varphi$ -code codes well on every  $R, Q$ -name in  $\Omega$ .

Given a positive number  $e$  less than  $\epsilon$ , we can pick a  $\varphi$ -code  $\Phi$  with  $\text{ErrorRate}(\Phi) < e/4$ . Set  $x \triangleq \Phi(\omega)$  and  $\check{x} \triangleq \Phi(\check{\omega})$ . Then  $\bar{d}(z, x) < e/4$  and  $\bar{d}(z, \check{x}) < e/4$  and therefore

$$(2.4) \quad \bar{d}(x, \check{x}) < e/2.$$

Recall that, in (1.7), the symbol  $\mathcal{C}$  denoted a code mapping the  $P$ -name  $x$  to  $\hat{x}$ , with error rate less than  $e$ . In our present context, in light of (2.4), we can just let  $\mathcal{C}$  be the identity code (codelength = 1) which maps the alphabet  $P$  to itself via the identity map.

*Emulating the setup in figure 2.* Let  $2l+1$  denote  $\text{CodeLen}(\Phi)$  and choose  $n$  sufficiently large that  $h' \gg l$ , where  $h'$  means  $h'_n$ . The  $Q$ - $h'$ -word  $W'$  codes, under  $\Phi$ , to a  $P$ - $(h'-2l)$ -word which we denote by  $W$ . Let  $h$  denote  $h'-2l$ . Agree to call  $x|_i^{i+h}$  a  $W$ -copy if it is the image of a  $W'$ -copy i.e. if  $\omega|_{-i}^{i+h+1}$  is a  $W'$ -copy. Similarly, say that  $\check{x}|_j^{j+h}$  is a  $W$ -copy if  $\check{\omega}|_{-j}^{j+h+1}$  is a  $W'$ -copy.

Now that we have defined the term  $W$ -copy, we can use figure 2 to define the gap between a  $W$ -copy and its successor  $W$ -copy, as well as the shift of a  $W$ -copy on  $x$  relative to the  $W$ -copy 'lying below' it on  $\check{x}$ . We can have chosen  $n$  sufficiently large that most  $W$ -copies on  $x$  have a shifted  $W$ -copy 'lying below'. Moreover, (2.3) implies that for most  $W$ -copies on  $x$ , the associated shift (as in figure 2) is non-zero (hence positive).

Now (2.2) implies there is a constant  $\sigma > 0$  such that the following version of the shift lemma, (1.4) holds. For any  $x$  and  $\check{x}$  obtained from a  $\Phi$  of sufficiently small error rate, and for any sufficiently large  $n$ : Suppose  $x|_i^{i+h}$  is a  $W$ -copy with positive shift, say  $s$ . Then  $s\% > \sigma$ .

*Mimicking the proof of (1.7).* Given  $\epsilon$ , use the periodicity extension lemma to get  $\epsilon'$ . In our factor context, this lemma is, in fact, easier to prove than in § 1. This is because our  $N$  codes will all be the same code – the identity code  $\mathcal{C}$ , which has a code length of 1. Pick some  $e$  so that  $2e/\sigma < \epsilon'$ . Then choose a  $\varphi$ -code  $\Phi$  with  $\text{ErrorRate}(\Phi) < e/4$  and set  $x \triangleq \Phi(\omega)$  and  $\check{x} \triangleq \Phi(\check{\omega})$ ; the role of the ' $N+1$  codes' from  $x$  to  $\check{x}$  will be played by the code  $\mathcal{C}$ . Pick  $\delta$  and  $\gamma$  and then choose  $n$  sufficiently large that  $h_n \gg \text{CodeLen}(\Phi)$ . Finally, say that a  $W$ -copy  $x|_i^{i+h}$  is a good  $W$ -copy if its shift is positive and if it satisfies the definition of good we used in (1.7.1).

Now we can replay the idea of the proof of (1.7) to conclude that  $\bar{d}(T^{-r}x, \check{x}) < \epsilon$  where  $r$  is the shift associated to some good  $W$ -copy and hence,  $r \neq 0$ . □

3. Given the two points  $\omega, \check{\omega} \in \Omega$  and a  $\delta > 0$ , say that the pair  $(\omega, \check{\omega})$   $\delta$ -match if, for infinitely many values of  $n$ : At least  $\delta$ -percent of  $\omega|_0^\infty$  is covered by  $W'$ -copies  $\omega|_i^{i+h'}$  for which  $\check{\omega}|_i^{i+h'}$  is also a  $W'$ -copy. In other words, the set of natural numbers

$$\{i: R^i(\omega), R^i(\check{\omega}) \in (\text{base of } n\text{-stack})\}$$

has lower density exceeding  $\delta/h'$ .



Our unproved fact of (2.3) was the existence of a pair  $(\omega, \tilde{\omega})$  which does not  $\delta$ -match (for any  $\delta$ , no matter how small) and such that  $\varphi(\omega) = \varphi(\tilde{\omega})$ . In order to state how to choose such a pair, we need the notion of the *relative independent joining* over a factor  $\mathcal{F}$ . So as to be self-contained, we will review the necessary facts about joinings.

We first need to recall a standard consequence of the ergodic theorem (lemma 3.2, below). We introduce some notation.

(3.1) *Definition.* Fix some invertible transformation  $\bar{R}$  on the space  $(\bar{\Omega}, \bar{\mathcal{A}}, \nu)$  with partition  $\bar{Q}$ ; denote the resulting stationary stochastic process by  $(\bar{R}, \bar{Q}; \nu)$ .

Suppose  $\eta$  and  $q$  are two  $\bar{R}, \bar{Q}$ -words of length  $l$  and  $k$ , respectively, where  $l > k$ . Index  $\eta$  from zero i.e.  $\eta = \eta|_0^l$ . Define the symbol freq ( $q$  on  $\eta$ ) to be the number

$$\frac{1}{l - (k - 1)} \cdot |\{i: 0 \leq i < i + k \leq l \text{ and } \eta|_i^{i+k} = q\}|.$$

Let  $\bar{Q}^k$  denote  $\bigvee_0^{k-1} \bar{R}^i \bar{Q}$ . The partition  $\bar{Q}^k$  is an ordered tuple of atoms, say

$$\bar{Q}^k = \langle q_1, \dots, q_m, \dots, q_M \rangle.$$

Define the symbol freq ( $\bar{Q}^k$  on  $\eta$ ) to be the probability vector (an  $M$ -tuple) whose  $m$ th component is freq ( $q_m$  on  $\eta$ ).

We write  $d_\nu(\bar{Q}^k)$ , the ‘distribution of the measure  $\nu$  on  $\bar{Q}^k$ ’, for the vector  $\langle \nu(q_1), \dots, \nu(q_M) \rangle$ . We remark in passing that this vector depends on  $\nu$ , whereas the previous vector does not.

Given two probability  $M$ -vectors  $\bar{a}$  and  $\bar{c}$ , let  $|\bar{a} - \bar{c}|$  denote the distance,  $\sum_{m=1}^M |a_m - c_m|$ , between them.

The lemma below is proved essentially as is the standard coding lemma (1.2).

(3.2) *LEMMA.* Suppose  $(\bar{R}, \bar{Q}; \nu)$  is an ergodic process. There is a set of  $\bar{R}, \bar{Q}$ -names  $\bar{\omega} \in \bar{\Omega}$  of full  $\nu$ -measure, such that for all positive  $\delta$  and  $\varepsilon$ ,  $\forall k, \forall$  large  $h'$ , the following holds: Cover, in any way, at least  $\delta$ -percent of  $\bar{\omega}|_0^\infty$  by disjoint blocks

$$\bar{\omega}|_{i_1}^{i_1+h'}, \bar{\omega}|_{i_2}^{i_2+h'}, \dots, \bar{\omega}|_{i_m}^{i_m+h'}, \dots.$$

Then at least  $\delta/2$ -percent of  $\bar{\omega}|_0^\infty$  is covered by those  $\bar{\omega}|_{i_m}^{i_m+h'}$  satisfying

$$(3.2.1) \quad |\text{freq}(\bar{Q}^k \text{ on } \bar{\omega}|_{i_m}^{i_m+h'}) - d_\nu(\bar{Q}^k)| < \varepsilon.$$

So, in particular, there exist values  $m$  such that (3.2.1) holds.

*Proof.* Follows from the ergodic theorem.

*Remark.* Agree to call a name  $\bar{\omega}$  generic for the process  $(\bar{R}, \bar{Q}; \nu)$ , if it is in the set of full  $\nu$ -measure of the lemma.

*Joinings.* Suppose  $(\Omega, \mathcal{A}, \mu)$  and  $(\Lambda, \mathcal{B}, m)$  are Lebesgue probability spaces. A measure  $\nu(\cdot)$  on  $(\Omega \times \Lambda, \mathcal{A} \times \mathcal{B})$  is called a *joining* of  $\mu$  and  $m$  if its marginal measures are  $\mu$  and  $m$ , respectively. That is,

$$\forall A \in \mathcal{A}: \nu(A \times \Lambda) = \mu(A)$$

and

$$\forall B \in \mathcal{B}: \nu(\Omega \times B) = m(B).$$

It is a technicality to check that a joining,  $\nu$ , of a Lebesgue measure with a Lebesgue measure must itself be a Lebesgue measure.

Now recall the transformation  $R: (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega, \mathcal{A}, \mu)$  of the preceding section. A *self-joining* (of  $R$ ) is a joining of  $\mu$  with  $\mu$  which is  $R \times R$  invariant. Self-joinings were introduced in [10]. Agree to call a self-joining  $\nu$  *ergodic* if the transformation

$$R \times R: (\Omega \times \Omega, \mathcal{A} \times \mathcal{A}, \nu) \rightarrow (\Omega \times \Omega, \mathcal{A} \times \mathcal{A}, \nu)$$

is ergodic. We abbreviate the above transformation by  $(R \times R; \Omega \times \Omega, \mathcal{A} \times \mathcal{A}, \nu)$ .

An example of a self-joining of  $R$  is *diagonal measure*  $\Delta(\cdot)$  which, like any joining, is completely determined once defined on rectangles:

$$\forall A, B \in \mathcal{A}: \Delta(A \times B) \triangleq \mu(A \cap B).$$

$\Delta$  is called diagonal measure because its support lies on  $\{(\omega, \omega): \omega \in \Omega\}$ , the diagonal of  $\Omega \times \Omega$ . Notice that  $(R \times R; \Omega \times \Omega, \mathcal{A} \times \mathcal{A}, \Delta)$  is isomorphic to  $(R; \Omega, \mathcal{A}, \mu)$  so  $\Delta$  is an ergodic self-joining.

Our factor algebra  $\mathcal{F}$  gives rise to another kind of self-joining. Define  $\rho_{\mathcal{F}}$ , the *relative independent joining* over  $\mathcal{F}$ , by:  $\forall A, B \in \mathcal{A}$

$$\rho_{\mathcal{F}}(A \times B) \triangleq \int_{\Omega} \mathbf{P}(A|\mathcal{F}) \cdot \mathbf{P}(B|\mathcal{F}) \, d\mu,$$

where  $\mathbf{P}(\cdot|\mathcal{F})$  denotes conditional probability. Notice that  $\rho_{\mathcal{F}}$  assigns full measure to the set  $\{(\omega, \check{\omega}): \varphi(\omega) = \varphi(\check{\omega})\}$  where  $\varphi$  is the factor homomorphism. Call this set the support of  $\rho_{\mathcal{F}}$ . Depending on the factor  $\mathcal{F}$ , the self-joining  $\rho_{\mathcal{F}}$  may or may not be ergodic. If it is not, it can be split into ergodic components.

(3.3) PROPOSITION. Any self-joining,  $\nu$ , has a decomposition of the following form:

$$(3.3.1) \quad \nu(D) = \int_J m_j(D) \, d\mathcal{J}(j),$$

for any set  $D \in \mathcal{A} \times \mathcal{A}$ . Here, each  $m_j(\cdot)$  is an ergodic self-joining of  $R$ , where  $j$  ranges over the ‘space of joinings’  $J$ . The measure  $\mathcal{J}(\cdot)$  is a probability measure on the space  $J$ .

*Proof.* Since  $\nu$  is a Lebesgue measure, we can apply the ergodic decomposition theorem to the transformation  $(R \times R; \Omega \times \Omega, \mathcal{A} \times \mathcal{A}, \nu)$ . It yields the decomposition (3.3.1) where, for each  $j$ ,  $(R \times R; \Omega \times \Omega, \mathcal{A} \times \mathcal{A}, m_j)$  is an ergodic transformation. So, we just need to show that each  $m_j$  is a self-joining by showing that its two marginals are each equal to  $\mu$ . Applying (3.3.1) to sets  $D$  of the form  $A \times \Omega$ , yields

$$(3.3.2) \quad \forall A \in \mathcal{A}: \mu(A) = \int_J m_j(A \times \Omega) \, d\mathcal{J}(j).$$

Treating  $m_j(\cdot \times \Omega)$  as a measure on  $(\Omega, \mathcal{A})$ , we see that (3.3.2) gives a decomposition of  $\mu$ , or rather, of  $(R; \Omega, \mathcal{A}, \mu)$ , into ergodic components. But the ergodic decomposition of a measure is unique and  $\mu$  is already ergodic! So, by discarding from  $J$  a set of  $j$  of  $\mathcal{J}$ -measure zero, we have that each  $m_j(\cdot \times \Omega)$  equals  $\mu(\cdot)$ . Repeating this argument for the other marginal yields the desired conclusion that each  $m_j$  is a self-joining. □

We can apply the above to  $\rho_{\mathcal{F}}$ . Since  $\mathcal{F}$  is a proper factor,  $\rho_{\mathcal{F}}$  is not diagonal measure. Consequently we can find some self-joining in the ergodic decomposition of  $\rho_{\mathcal{F}}$

which is not diagonal measure. This self-joining assigns full measure to the support of  $\rho_{\mathcal{F}}$ .

(3.4) PROPOSITION.  $\exists$  an ergodic self-joining  $\rho$ ,  $\rho \neq \Delta$ , such that

$$\rho(\{(\omega, \check{\omega}) : \varphi(\omega) = \varphi(\check{\omega})\}) = 1.$$

(3.5) LEMMA. Suppose  $\nu$  is an ergodic self-joining and  $(\omega, \check{\omega})$  is generic, in the sense of lemma 3.2, for the process  $(R \times R, Q \times Q; \nu)$ . If  $\exists \delta > 0$  such that  $(\omega, \check{\omega})$   $\delta$ -match, then  $\nu = \Delta$ .

*Proof.* It suffices to show that  $\nu = \Delta$  on the algebra of rectangles of the form  $R, Q$ -cylinder-set cross  $R, Q$ -cylinder-set. So, given a length  $k$  and an  $\epsilon$ , we need to show

$$|d_\nu(\bar{Q}^k) - d_\Delta(\bar{Q}^k)| < 2\epsilon,$$

where  $\bar{Q}^k$  is to denote the partition  $\bigvee_0^{k-1} (R \times R)^i (Q \times Q)$ .

Since  $(\omega, \check{\omega})$   $\delta$ -match, the following holds for infinitely many  $n$  and hence, without loss of generality, for all large  $n$  (by having dropped to a subsequence and renumbered):

$\exists$  indices  $0 \leq i_1 \leq i_2 < \dots < i_m < \dots$  such that for each  $m$ ,  $\omega|_{i_m}^{i_m+h'}$  and  $\check{\omega}|_{i_m}^{i_m+h'}$  are  $W'$ -copies.

Moreover, the blocks  $\{(\omega, \check{\omega})|_{i_m}^{i_m+h'} : m = 1, 2, \dots\}$  cover at least  $\delta$ -percent of  $(\omega, \check{\omega})|_0^\infty$ . Consequently lemma 3.2 asserts, our having chosen  $n$  adequately large, that at least one of these blocks is pretty typical of the measure  $\nu$ . But these blocks are all identical and equal to  $(W'_n, W'_n)$ . Thus

$$(3.5.1) \quad \forall \text{ large } n: |\text{freq}(\bar{Q}^k \text{ on } (W'_n, W'_n)) - d_\nu(\bar{Q}^k)| < \epsilon.$$

Let us now derive the same statement with  $\nu$  replaced by  $\Delta$ . For any  $\omega' \in \Omega$ , the pair  $(\omega', \omega')$  is generic for  $(R \times R, Q \times Q; \Delta)$ . In particular, the pair  $(\omega, \omega)$  is generic. The blocks  $\{(\omega, \omega)|_{i_m}^{i_m+h'} : m = 1, 2, \dots\}$  cover at least  $\delta$ -percent of  $(\omega, \omega)|_0^\infty$ . So, as before, lemma 3.2 implies:

$$\forall \text{ large } n: |\text{freq}(\bar{Q}^k \text{ on } (W'_n, W'_n)) - d_\Delta(\bar{Q}^k)| < \epsilon.$$

This together with (3.5.1) yields  $|d_\nu(\bar{Q}^k) - d_\Delta(\bar{Q}^k)| < 2\epsilon$  as desired. □

Now we know how to pick a  $(\omega, \check{\omega})$  that never  $\delta$ -match and such that  $\varphi(\omega) = \varphi(\check{\omega})$ . By proposition 3.4 we can pick  $(\omega, \check{\omega})$  generic for the joining  $\rho$  and such that  $\varphi(\omega) = \varphi(\check{\omega})$ . Since  $\rho \neq \Delta$ , lemma 3.5 ensures that  $(\omega, \check{\omega})$  do not  $\delta$ -match for any  $\delta$ .

We now know how to choose the pair  $(\omega, \check{\omega})$  that we needed in (2.3). This completes the proof that the factor transformation is rigid.

*Remark.* A restatement of the rigidity of factors is

(\*) A non-rigid transformation cannot sit as a (proper) factor of a rank-1 transformation.

In particular, this is true for the rank-1 transformations of Ornstein and of Chacón (examples (i) and (iii) of §0) since they are non-rigid.

Ornstein's transformation,  $T$ , is prime – it has no factor. Actually, since any factor of a rank-1 must be rank-1, it is equivalent to say that  $T$  is prime in the class rank-1, that is, it has no rank-1 factor. In this light, (\*) says that  $T$  is 'reverse' prime, in

the class rank-1. The transformation of example (iv) in § 0, is a transformation which is reverse prime but not (quite) prime.

*Open question.* For a fixed rank-1  $T: (X, \mu) \rightarrow (X, \mu)$  and for each  $S \in C(T)$ , one can define a self-joining of  $T$  whose support lies on the graph of  $S$ . This ‘graph measure’,  $\nu_S$ , is defined on rectangles by

$$\nu_S(A \times B) \triangleq \mu(A \cap S^{-1}(B)).$$

In particular, for each  $k \in \mathbb{Z}$  we get a joining  $\nu_{T^k}$  called an off-diagonal joining; denote it by  $\Delta^k$ . The weak-closure theorem, restated, is that any graph joining  $\nu_S$  is a weak limit (in the sense of measures) of off-diagonal joinings. That is, we have a sequence  $\{n_i\}$  such that

$$\forall A, B: |\Delta^{n_i}(A \times B) - \nu_S(A \times B)| \rightarrow_i 0.$$

This suggests an intriguing question: Must every ergodic self-joining be a weak limit of off-diagonals? This is true for product measure  $\rho(A \times B) \triangleq \mu(A) \cdot \mu(B)$  since

$$\begin{aligned} \exists \{n_i\}: \Delta^{n_i} \rightarrow \rho &\Leftrightarrow \exists \{n_i\}, \forall A, B: \mu(A \cap T^{-n_i}B) \rightarrow \mu(A)\mu(B) \\ &\Leftrightarrow T \text{ is weak mixing} \\ &\Leftrightarrow \rho \text{ is ergodic.} \end{aligned}$$

*Three remarks.* (In this section, all transformations are invertible.)

With the notion of self-joinings we can elaborate a bit on corollaries (A) and (C) of the weak-closure theorem.

Suppose  $S$  is a weak limit  $S = \lim_{i \rightarrow \infty} T^{n_i}$  of powers of a transformation  $T$  on space  $(X, \mathcal{A}, \mu)$ . Then  $S$  inherits the commutant of  $T$  i.e.  $C(S) \supset C(T)$ . Also,  $S$  inherits as factor algebras, all the factor algebras of  $T$ . More generally, any self-joining of  $T$  is a self-joining for  $S$ . For if a measure  $\nu(\cdot)$  on  $(X \times X, \mathcal{A} \times \mathcal{A})$  is  $T \times T$ -invariant then it is  $T^{n_i} \times T^{n_i}$ -invariant; hence it is  $S \times S$ -invariant since

$$\begin{aligned} &\nu((S \times S)^{-1}(A \times B) \Delta (T \times T)^{-n_i}(A \times B)) \\ &\leq \mu(S^{-1}A \Delta T^{-n_i}(A)) + \mu(S^{-1}B \Delta T^{-n_i}(B)), \end{aligned}$$

where the two latter terms go to zero as  $i \rightarrow \infty$ . (This inequality follows from the containment relation

$$(A \times B) \Delta (A' \times B') \subset [(A \Delta A') \times X] \cup [X \times (B \Delta B')],$$

which holds for any four sets  $A, B, A', B' \subset X$ .) We see thus that  $S$  inherits all of the self-joinings of  $T$ .

It is worth noting that if  $S$  turns out to be rank-1, then since  $T \in C(S) = \text{WCl}(S)$ ,  $T$  inherits these properties from  $S$ . Hence  $T$  and  $S$  have identical commutant, factor algebras, and self-joinings. For example, not surprisingly, all irrational rotations on the unit circle share these properties. As another instance, the weak mixing  $T$  of example (v), in § 0, has all its non-zero powers rank-1; so these powers  $T^k$  have common commutant, factor algebras, and joinings.

A comment in passing: The ‘weak-limit relation’ is a transitive relation i.e. for transformations  $S_1, S_2$ , and  $S_3$  on the same space,

$$[S_2 \in \text{WCl}(S_1) \text{ and } S_3 \in \text{WCl}(S_2)] \Rightarrow S_3 \in \text{WCl}(S_1).$$

The ‘commuting relation’ is trivially reflexive ( $S_1$  commutes with  $S_2 \Rightarrow S_2$  commutes with  $S_1$ ). In the class of rank-1 transformations, the ‘weak-limit relation’ and the ‘commuting relation’ are the same relation; hence it is an equivalence relation. This relation partitions the class Rank-1 into equivalence classes. Either every element of such an equivalence class is weak mixing, or no element is; this follows because (rotation) factors are inherited under weak limits. A ‘weak mixing’ equivalence class can be uncountable – a future paper will give an example of an equivalence class containing an uncountable abelian group (except the identity element, which is not rank-1) of weak-mixing rank-1 transformations.

*Remark 2.* With the notion of relative independent joining, we can refine corollary (C) of the weak-closure theorem. The corollary says that a rank-1 transformation – say, on a space  $(X, \mathcal{A}, \mu)$ , can have no factor which is isomorphic to a cartesian square. In other words, if factor algebras  $\mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{A}$  yield isomorphic factor transformations then  $\mathcal{F}_1$  and  $\mathcal{F}_2$  cannot sit independently (orthogonally) within  $\mathcal{A}$ .

A bit more is true. They cannot, in fact, sit independently even relative to a non-trivial common factor. In other words, no factor of a rank-1 transformation can arise as a relative independent joining. Since any factor of rank-1 is itself rank-1, it suffices to show the following.

*Corollary (C) revisited.* Suppose we have a transformation  $R: (Y, \mathcal{A}, \mu) \rightarrow (Y, \mathcal{A}, \mu)$  and a factor algebra  $\mathcal{F} \subset \mathcal{A}$ . Let  $\rho (= \rho_{\mathcal{F}})$  denote the relative independent joining measure. Let  $T$  denote the transformation

$$R \times R: (Y \times Y, \mathcal{A} \times \mathcal{A}, \rho) \rightarrow (Y \times Y, \mathcal{A} \times \mathcal{A}, \rho).$$

Then  $T$  cannot be rank-1.

*Proof.* Let  $S$  be the coordinate flip map on  $Y \times Y$  which sends  $(y, y') \mapsto (y', y)$ .  $S$  is measurable and commutes with  $R \times R$ . Moreover,

$$\rho(S^{-1}(A \times B)) = \rho(B \times A) \triangleq \int_Y \mathbf{P}(B | \mathcal{F}) \cdot \mathbf{P}(A | \mathcal{F}) \, d\mu = \rho(A \times B).$$

$S$  preserves the measure  $\rho$  so  $S \in C(T)$ . Now if  $T$  is rank-1 then  $S$  must leave invariant any factor algebra of  $T$ ; in particular, the factor algebra  $\{\phi, Y\} \times \mathcal{A}$ . For any  $A \in \mathcal{A}$  the set  $S^{-1}(Y \times A)$  must be in that same algebra. So there is a set  $Y \times B$  in that algebra which equals  $A \times Y$ ,  $\rho$ -a.e. Thus  $\mu(B) = \mu(A)$  and

$$\rho(Y \times B) = \rho([A \times Y] \cap [Y \times B]) = \rho(A \times B).$$

Hence

$$0 = \rho(Y \times B) - \rho(A \times B) = \int_Y [1 - \mathbf{P}(A | \mathcal{F})] \cdot \mathbf{P}(B | \mathcal{F}) \, d\mu.$$

The integrand is a product of two non-negative functions, so  $\mathbf{P}(A | \mathcal{F})$  must be identically 1 on the set  $\{\mathbf{P}(B | \mathcal{F}) > 0\}$ . Since  $\mathbf{P}(B | \mathcal{F}) \leq 1 = \mathbf{P}(A | \mathcal{F})$  on this set and  $\mu(B) = \mu(A)$ , we see that  $\mathbf{P}(A | \mathcal{F})$  can be none other than the indicator function of this set. Hence  $A \in \mathcal{F}$ . We are forced to conclude that  $\mathcal{F} = \mathcal{A}$ . □

We mention, without proof, that joint work with Dan Rudolph yields the stronger statement that rank-1 transformations have unique factors. If  $T$  is rank-1 with factor

algebras  $\mathcal{F}$  and  $\mathcal{G}$ , then

$$T|_{\mathcal{F}} \cong T|_{\mathcal{G}} \Rightarrow \mathcal{F} = \mathcal{G}.$$

*Remark 3.* This last comment also concerns corollary (C) of the weak closure theorem. We know that if a rank-1  $T$  is a cartesian product  $R_1 \times R_2$ , where each  $(R_i; Y_i, \mathcal{A}_i, \mu_i)$  is a transformation, then  $R_1$  and  $R_2$  are relatively prime. Somewhat stronger, they must be *disjoint*, in the sense of Furstenberg [3]. (A *joining* of a transformation  $R_1$  with a transformation  $R_2$  is an  $R_1 \times R_2$ -invariant measure on  $(Y_1 \times Y_2, \mathcal{A}_1 \times \mathcal{A}_2)$  possessing marginals  $\mu_1$  and  $\mu_2$ , respectively. Transformations  $R_1$  and  $R_2$  are *disjoint* if the only joining of them is product measure  $\mu_1 \times \mu_2$ .)

Suppose  $\nu(\cdot)$  is a joining of  $R_1$  with  $R_2$  where  $R_1 \times R_2$  is rank-1. The transformation  $R_1 \times \text{Id}_2$  on  $Y_1 \times Y_2$  commutes with  $R_1 \times R_2$ ; so  $R_1 \times \text{Id}_2 \in \text{WCl}(R_1 \times R_2)$ . Consequently, by the same argument that shows that self-joinings are inherited under weak-limits,  $\nu$  is an  $R_1 \times \text{Id}_2$ -invariant measure. This together with the ergodicity of  $R_1$  implies that  $\nu = \mu_1 \times \mu_2$ , by proposition 2 of [11].

This twofold disjointness coupled with an induction argument, imply that if a countable cartesian product  $R_1 \times R_2 \times \dots$  is rank-1, then the  $\{R_i\}$  are mutually disjoint. This means that if  $\nu(\cdot)$  is an  $(R_1 \times R_2 \times \dots)$ -invariant measure on the product space, with respective marginals  $\mu_1(\cdot), \mu_2(\cdot), \dots$ , then in fact  $\nu$  is product measure

$$\nu = \mu_1 \times \mu_2 \times \dots.$$

This work was done during a most pleasant stay at the University of Toronto during the summer of 1984; I am grateful to Professor Akcoglu for giving me the opportunity to visit. My thanks also to his doctoral student Chris Bose, for many thought-provoking conversations and for willingly suffering early vague versions of this proof.

The writing of this paper was partially supported by grant # DMS85 0 1519.

#### REFERENCES

- [1] M. A. Akcoglu, R. V. Chacón & T. Schwartzbauer. Commuting transformations and mixing. *Proc. Amer. Math. Soc.* **24** (1970), 637–642, MR 40 #7421.
- [2] N. Friedman & D. S. Ornstein. On partially mixing transformations. *J. Math. Mech.* **20** (1971), 767–775.
- [3] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory* **1** (1967), 1–49.
- [4] A. del Junco. A simple measure-preserving transformation with trivial centralizer. *Israel J. Math.* **14** (1973), 26–38.
- [5] A. del Junco. Transformations with discrete spectrum are stacking transformations. *Can. J. Math.* **28** #4 (1976), 836–839.
- [6] A. del Junco, M. Rahe & L. Swanson. Chacón's automorphism has minimal self-joinings. *J. Analyse Math.* **37** (1980), 276–284.
- [7] A. B. Katok, Ya. G. Sinai & A. M. Stepin. Theory of dynamical systems and general transformation groups with invariant measure. *J. Sov. Math.* **7** (1977), 974–1065.
- [8] J. King. Remarks on the commutant and factor algebras of finite rank mixing transformations. In preparation.
- [9] D. S. Ornstein. On the root problem in ergodic theory. In *Proc. Sixth Berkeley Symp. Math. Stat. Prob. Vol. II*, 347–356. Univ. of California Press (1967).
- [10] D. Rudolph. An example of a measure-preserving map with minimal self-joinings. *J. d'Analyse Math.* **35** (1979), 97–122.
- [11] A. del Junco & D. Rudolph. On ergodic actions whose self-joinings are graphs Preprint.