Mixing of All Orders and Pointwise Averages

Bogdan Grechuk · 10 Nov 2026

https://functor.network/user/3333/entry/2061

In September and October 2026, OpenAI released two preprints (OpenAI 2026b, 2026a) concerning fundamental problems in ergodic theory. The first announces a proof of Rokhlin's multiple-mixing conjecture, which had remained open for more than seventy years. The second establishes pointwise convergence of multiple ergodic averages for mixing transformations.

The two results are closely related, so we discuss them together.

Let us recall some definitions.

A *measure-preserving dynamical system* consists of a probability space $(X, \mathcal{B}, \mu)$ and a measurable transformation $T : X \to X$ such that

$$\mu(T^{-1}A) = \mu(A)$$

for every measurable set $A \subseteq X$. Here $\mu(X) = 1$, and the equality means that the probability of an event does not change when the system evolves by one time step.

We assume that T is invertible and that its inverse is measurable. Starting from a point $x \in X$, we obtain its *orbit*

$$x,\ T(x),\ T^2(x),\ T^3(x), \dots,$$

where T^n means that T is applied n times.

One of the basic questions in ergodic theory is whether the system gradually loses information about its initial state.

A measure-preserving transformation is called *mixing* if, for every pair of measurable sets $A, B \subseteq X$,

$$\lim_{n\to\infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

The interpretation is simple. Choose the initial state x at random according to μ. The probability that x belongs to A and that its state after n steps belongs to B is

$$\mu(A \cap T^{-n} B).$$

If T is mixing, this probability approaches $\mu(A)\mu(B)$, as if the two observations were independent.

A standard example is the *Bernoulli shift*. Imagine an infinite sequence of independent fair coin tosses, indexed by all integers, and let T shift the sequence one position to the left. Events determined by finite blocks of coin tosses become independent once their blocks are sufficiently far apart. This leads to the mixing property.

But what happens if we observe the same system at *three or more* widely separated times?

For example, suppose that A, B, C are measurable sets. Does ordinary mixing imply that

$$\mu(A \cap T^{-n} B \cap T^{-2n} C) \longrightarrow \mu(A)\mu(B)\mu(C)$$

as $n \to \infty$?

The difficulty is that pairwise independence does not necessarily imply joint independence.

For example, let U, V be independent random variables, each taking the values 0 and 1 with equal probability, and put

$$W = U + V \pmod 2.$$

Every pair among U, V, W is independent. However, knowing U and V determines W completely. In particular,

$$\mathbb{P}(U = V = W = 0) = \frac{1}{4},$$

whereas the product of their individual probabilities is $1/8$.

Thus, knowing that pairs of observations become independent does not immediately tell us what happens to larger collections.

In 1949, Rokhlin raised the question of whether a mixing transformation must also be *mixing of all orders*. This means that, for every integer $r \geq 2$ and measurable sets $A_1, \ldots, A_r$, we should have

$$\mu\left(\bigcap_{j=1}^{r} T^{-n_j} A_j\right) \longrightarrow \prod_{j=1}^{r} \mu(A_j)$$

whenever all the pairwise differences $|n_i - n_j|$ tend to infinity.

The case $r = 2$ is precisely the definition of ordinary mixing.

Over the decades, the conjecture was established for several important classes of dynamical systems. For example, Kalikow (Kalikow 1984) proved that mixing implies threefold mixing for rank-one transformations. Ryzhikov (Ryzhikov 1993) subsequently obtained mixing of all orders for mixing finite-rank transformations.

However, the problem for arbitrary mixing transformations remained open.

There was also a reason to be cautious. In 1978, Ledrappier (Ledrappier 1978) constructed an action of two commuting transformations that is mixing in the appropriate twofold sense but not threefold mixing. Thus, the analogous statement fails for certain actions of the group $\mathbb{Z}^2$.

The crucial question was whether this could happen for the iterates of a *single* transformation.

The first new preprint (OpenAI 2026b), posted on September 23, 2026, gives an affirmative answer to Rokhlin's conjecture.

Theorem 1 (OpenAI) *Every invertible mixing probability-preserving transformation is mixing of every finite order.*
More precisely, for every integer $r \geq 2$ and all measurable sets $A_1, \ldots, A_r \subseteq X$,

$$\mu\left(\bigcap_{j=1}^{r} T^{-n_j} A_j\right) \longrightarrow \prod_{j=1}^{r} \mu(A_j)$$

as

$$\min_{i \neq j} |n_i - n_j| \longrightarrow \infty.$$

Thus, for a single invertible transformation, the asymptotic independence of pairs of observations already forces asymptotic independence of every fixed finite number of observations.

Let us briefly discuss the proof.

The authors argue by contradiction, supposing that mixing holds but fails at some higher order. They choose the smallest order at which this happens and study the limiting joint distributions of observations taken at increasingly separated times.

These limiting distributions are called *joinings*. They retain information about possible correlations between the observations even when all smaller collections have become independent.

The proof constructs increasingly elaborate arrays of such joinings and analyzes the conditional expectations connecting their coordinates. This produces algebraic restrictions on any correlations that might survive at widely separated times.

Finally, a cancellation argument, using ordinary mixing and information-theoretic properties of the system, rules out the remaining correlations. This contradicts the assumed failure of higher-order mixing.

The second new theorem concerns a related but different question.

For bounded measurable functions $g_1, \dots, g_r : X \to \mathbb{R}$, consider the *multiple ergodic averages*

$$A_N(x) = \frac{1}{N} \sum_{k=1}^{N} g_1(T^k x) g_2(T^{2k} x) \cdots g_r(T^{rk} x).$$

These averages combine observations of the same orbit at the times

$$k, 2k, \dots, rk.$$

When $r = 1$, we obtain the ordinary ergodic averages

$$\frac{1}{N} \sum_{k=1}^{N} g_1(T^k x),$$

whose pointwise convergence almost everywhere follows from Birkhoff's classical ergodic theorem.

When $r \geq 2$, the problem becomes much more difficult. In 1990, Bourgain (Bourgain 1990) proved pointwise convergence for the case of two linear iterates. Host and Kra (Host and Kra 2005), and independently Ziegler, subsequently established convergence in the L^2 norm for any finite number of consecutive linear iterates.

However, convergence in L^2 does not automatically imply pointwise convergence along the entire sequence of positive integers. The latter requires control of the averages for almost every individual starting point.

In a previous blog post, we discussed a related theorem of Kosz, Mirek, Peluse, Wan and Wright concerning pointwise convergence of polynomial multiple averages when the polynomials have distinct degrees.

The new result treats a complementary situation. All the polynomials are linear,

$$k, 2k, \dots, rk,$$

but the transformation is assumed to be mixing.

The preprint (OpenAI 2026a), posted on October 4, 2026, establishes the following theorem.

Theorem 2 (OpenAI) *Let T be an invertible mixing probability-preserving transformation. Then, for every integer $r \geq 2$ and every fixed collection of bounded measurable functions $g_1, \ldots, g_r : X \to \mathbb{R}$,*

$$\frac{1}{N} \sum_{k=1}^{N} \prod_{j=1}^{r} g_j(T^{jk}x) \longrightarrow \prod_{j=1}^{r} \int_X g_j \, d\mu$$

for almost every $x \in X$, as $N \to \infty$.
For example, taking g_1, g_2, g_3 to be the indicator functions of measurable sets A, B, C, respectively, gives

$$\lim_{N\to\infty} \frac{1}{N} \#\{1 \leq k \leq N : T^k x \in A,\ T^{2k}x \in B,\ T^{3k}x \in C\} = \mu(A)\mu(B)\mu(C)$$

for almost every starting point x.

Thus, although the observations at times $k, 2k, 3k$ are made along the same orbit, their simultaneous occurrence has the expected asymptotic frequency.

It is worth emphasizing the distinction between the two new theorems.

The first theorem implies that, for bounded functions $g_1, \ldots, g_r$,

$$\int_X \prod_{j=1}^{r} g_j(T^{jk}x) \, d\mu(x) \longrightarrow \prod_{j=1}^{r} \int_X g_j \, d\mu$$

as $k \to \infty$.

But this is a statement about the *average over all starting points*. It does not by itself establish the pointwise convergence of $A_N(x)$ for almost every individual x.

The second theorem requires substantial additional work. Its proof combines higher-order mixing with pointwise estimates for triple averages and a sophisticated analysis of limiting distributions of multiple observations. By studying hypothetical failures of pointwise convergence, the authors construct special joinings and use their structure to obtain a contradiction.

Neither theorem assumes a quantitative rate of mixing. In particular, the conclusions apply even when the correlations in the definition of mixing converge to their limits extremely slowly.

Both papers are currently available as preprints. If their proofs withstand independent verification, they will settle two major questions about the statistical behaviour of deterministic dynamical systems.

Together, the results give a powerful principle: for an invertible mixing transformation, ordinary asymptotic independence of pairs of observations forces not only mixing of every finite order, but also the expected pointwise frequencies of simultaneous observations along arbitrarily long arithmetic progressions of times.

References

Bourgain, Jean. 1990. “Double Recurrence and Almost Sure Convergence.” *Journal für Die Reine Und Angewandte Mathematik* 404: 140–61. https://doi.org/10.1515/crll.1990.404.140.

Host, Bernard, and Bryna Kra. 2005. “Nonconventional Ergodic Averages and Nilmanifolds.” *Annals of Mathematics*, 2nd series, vol. 161 (1): 397–488.

Kalikow, Steven Arthur. 1984. “Twofold Mixing Implies Threefold Mixing for Rank One Transformations.” *Ergodic Theory and Dynamical Systems* 4 (2): 237–59. https://doi.org/10.1017/S014338570000242X.

Ledrappier, François. 1978. “Un Champ Markovien Peut être d’entropie Nulle Et mélangeant.” *Comptes Rendus de l’Académie Des Sciences de Paris, Série A–B* 287 (7): A561–63.

OpenAI. 2026a. *Pointwise Multiple Ergodic Averages for Mixing Transformations*. Preprint. https://github.com/openai/math/blob/main/preprints/Pointwise-Multiple-Ergodic-Averages-for-Mixing-Transformations-October-4-2026/multiple-ergodic-averages.pdf.

OpenAI. 2026b. *Rokhlin’s Multiple-Mixing Problem for One Transformation*. Preprint. https://github.com/openai/math/blob/main/preprints/Rokhlins-multiple-mixing-problem-for-one-transformation-September-23-2026/paper.pdf.

Ryzhikov, V. V. 1993. “Joinings and Multiple Mixing of Finite Rank Actions.” *Functional Analysis and Its Applications* 27 (2): 128–40. https://doi.org/10.1007/BF01085983.