Non-Abelian eigenstate thermalization hypothesis

Chaitanya Murthy,[1] Arman Babakhani,[2,3] Fernando Iniguez,[4] Mark Srednicki,[4] and Nicole Yunger Halpern[5,6,*]

[1] *Department of Physics, Stanford University, Stanford, CA 94305, USA*
[2] *Department of Physics, University of Southern California, Los Angeles, CA, 90089, USA*
[3] *Information Sciences Institute, Marina Del Rey, CA, 90292, USA*
[4] *Department of Physics, University of California, Santa Barbara, CA 93106, USA*
[5] *Joint Center for Quantum Information and Computer Science, NIST and University of Maryland, College Park, MD 20742, USA*
[6] *Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA*

(Dated: June 14, 2022)

The eigenstate thermalization hypothesis (ETH) explains why chaotic quantum many-body systems thermalize internally if the Hamiltonian lacks symmetries. If the Hamiltonian conserves one quantity ("charge"), the ETH implies thermalization within a charge sector—in a microcanonical subspace. But quantum systems can have charges that fail to commute with each other and so share no eigenbasis; microcanonical subspaces may not exist. Furthermore, the Hamiltonian will have degeneracies, so the ETH need not imply thermalization. We adapt the ETH to noncommuting charges by positing a *non-Abelian ETH* and invoking the *approximate microcanonical subspace* introduced in quantum thermodynamics. Illustrating with SU(2) symmetry, we apply the non-Abelian ETH in calculating local observables' time-averaged and thermal expectation values. In many cases, we prove, the time average thermalizes. However, we also find cases in which, under a physically reasonable assumption, the time average converges to the thermal average unusually slowly as a function of the global-system size. This work extends the ETH, a cornerstone of many-body physics, to noncommuting charges, recently a subject of intense activity in quantum thermodynamics.

"Chaotic" closed quantum many-body systems thermalize internally, in the absence of conserved charges: Typical observables $\mathscr{O}$ equilibrate to the expectation values they would have in the canonical state $\rho_{\rm can} \propto e^{-\beta H}$. H denotes the Hamiltonian, whose expectation value determines the inverse temperature β [1]. The eigenstate thermalization hypothesis (ETH) explains this thermalization [2–4]: Let $|\alpha\rangle$ denote the energy eigenstates; E_α, the eigenenergies; and $\mathscr{O}_{\alpha\alpha'} \coloneqq \langle\alpha|\mathscr{O}|\alpha'\rangle$, matrix elements representing the observable. $\mathscr{O}$ and H satisfy the ETH if

$$\mathscr{O}_{\alpha\alpha'} = \mathcal{O}(\mathcal{E})\,\delta_{\alpha,\alpha'} + e^{-S_{\rm th}(\mathcal{E})/2}\,f(\mathcal{E},\omega)R_{\alpha\alpha'}\,. \qquad (1)$$

The two energies' average is $\mathcal{E} \coloneqq (E_\alpha + E_{\alpha'})/2$; their difference is $\omega \coloneqq E_\alpha - E_{\alpha'}$; $\mathcal{O}(\mathcal{E})$ and $f(\mathcal{E},\omega)$ are smooth, real functions; $S_{\rm th}(\mathcal{E})$ denotes the thermodynamic entropy (the logarithm of the density of states) at the energy $\mathcal{E}$; $\delta_{\alpha,\alpha'}$ denotes the Kronecker delta; and R is a Hermitian matrix of erratically varying elements, which have zero means and unit variances across small E_α and $E_{\alpha'}$ windows [5, 6]. The first, "diagonal" (α=α') term in Eq. (1) contains the microcanonical expectation value $\mathcal{O}(\mathcal{E})$. The thermodynamic entropy $S_{\rm th}(\mathcal{E})$ exponentially suppresses the second, "off-diagonal" term. The ETH explains behaviors observed numerically and experimentally in condensed matter; atomic, molecular, and optical physics; and high-energy physics (e.g., [1, 7–16]).

If $\mathscr{O}$ and a nondegenerate H satisfy the ETH, $\mathscr{O}$ thermalizes [1, 17], in the following sense. Let N denote the size of the global system, which begins in a normalized state $|\psi(0)\rangle = \sum_\alpha C_\alpha|\alpha\rangle$ with an extensive energy $E \coloneqq \langle H\rangle = O(N)$.[1] Let the energy variance, $\mathrm{var}(H) \coloneqq \langle H^2\rangle - E^2$, be $\leq O(N)$. At a time t, the observable has an expectation value

$$\langle\mathscr{O}\rangle_t = \sum_\alpha |C_\alpha|^2\mathscr{O}_{\alpha\alpha} + \sum_{\alpha\neq\alpha'} C_\alpha^* C_{\alpha'}\, e^{i(E_\alpha - E_{\alpha'})t/\hbar}\,\mathscr{O}_{\alpha\alpha'}\,. \qquad (2)$$

Consider averaging this value over an infinite time: $\overline{\langle\mathscr{O}\rangle}_t \coloneqq \lim_{t\to\infty}\frac{1}{t}\int_0^t dt'\,\langle\mathscr{O}\rangle_{t'}$. As H lacks degeneracies, phase cancellations ensure that the second term averages to zero: $\overline{\langle\mathscr{O}\rangle}_t = \sum_\alpha |C_\alpha|^2\mathscr{O}_{\alpha\alpha}$. By the energy-variance bound, $|C_\alpha|$ is significant only for α values in a narrow window. By the ETH (1), $\mathscr{O}_{\alpha\alpha} \approx \mathcal{O}(\mathcal{E})$ can approximately be factored out of the sum, which then equals one by the state's normalization. Therefore, the time-averaged expectation value $\overline{\langle\mathscr{O}\rangle}_t$ approximately equals the microcanonical average $\mathcal{O}(\mathcal{E})$ [18]. The latter approximates the canonical average, $\mathrm{Tr}(\mathscr{O}\rho_{\rm can})$, by the thermodynamic ensembles' equivalence at large N.

At all times, the difference $\langle\mathscr{O}\rangle_t - \overline{\langle\mathscr{O}\rangle}_t$ depends on the second term in (2). This term is typically exponentially small in N, by the off-diagonal term in the ETH (1). This exponential smallness can be overcome at an instant (e.g., $t = 0$), if exponentially many C_α's have relative phases such that the $C_\alpha^* C_{\alpha'} R_{\alpha\alpha'}$'s add coherently, overpowering the $e^{-S_{\rm th}(\mathcal{E})/2}$. But time evolution destroys this coherence, and $\langle\mathscr{O}\rangle_t$ relaxes to $\overline{\langle\mathscr{O}\rangle}_t \approx \mathrm{Tr}(\mathscr{O}\rho_{\rm can})$.

* nicoleyh@umd.edu

[1] We use big-O notation as in many-body physics, meaning "scales as," rather than to encode an upper bound, as in information theory.

This argument for thermalization relies on the Hamiltonian's nondegeneracy and the matrix-element structure (1). Both postulates are questionable if H conserves charges.[2] If H has an Abelian symmetry, the energy spectrum can lack degeneracies. Since the charges commute, they share eigenspaces—charge sectors. In each shared charge sector, the ETH applies. For example, many spin-system dynamics conserve the total spin's z-component, S_z. Similarly, many particle dynamics conserve total particle number. The ETH is often applied in an S_z sector or a particle-number sector, wherein Eq. (1) and the arguments above hold.

A non-Abelian symmetry can eliminate our recourse to a charge sector: Such a symmetry is generated by charges Q_a that fail to commute with each other and so cannot necessarily be assigned definite values simultaneously. Furthermore, non-Abelian symmetries have irreducible representations of dimension > 1, generically forcing degeneracies on H. Finally, in opposition to (1), the matrix elements $\mathscr{O}_{\alpha\alpha'}$ are constrained by the observable's transformations under the symmetry operations.

For example, consider N qubits (quantum two-level systems). H can conserve the global spin components $S_{a=x,y,z}$, inducing an SU(2) symmetry. The energy spectrum splits into degenerate multiplets labeled by total spin quantum numbers s_α. Only the singlets, whose $s_\alpha = 0$, are simultaneous eigenspaces of all the S_a. The matrix elements $\mathscr{O}_{\alpha\alpha'}$ obey the Wigner-Eckart theorem [Eq. (13)], conflicting with the naïve ETH (1).

We bridge the ETH to the thermodynamics of noncommuting charges, a subfield emerging at the intersection of quantum information and quantum thermodynamics [19–49]. We propose a non-Abelian ETH and apply it to SU(2) symmetry [Eq. (14)], expecting the results to generalize. Using the non-Abelian ETH, we compute two averages of K-local observables: time-averaged and thermal expectation values. For many observables and initial states, we find, the time average agrees with the thermal average: Deviations are $O(N^{-1})$, as without noncommuting charges [50]. For certain observables and $|\psi(0)\rangle$'s, however, the time average deviates from the thermal prediction by anomalously large corrections $\sim N^{-1/2}$. This result holds under a physically reasonable assumption about the non-Abelian analog of $\mathscr{O}_{\alpha\alpha'}$. Our argument employs the Wigner–Eckart theorem; properties of Clebsch–Gordan coefficients; and approximate microcanonical subspaces, which generalize microcanonical subspaces to accommodate noncommuting charges [22, 35, 48]. This work extends the ETH, a mainstay of many-body physics, to the more fully quantum domain of noncommuting charges and so to quantum-information thermodynamics. Below, we introduce our setup, present the non-Abelian ETH, apply it to calculate observables' thermal and time-averaged expectation values, compare the averages, and outline the opportunities established by this work.

Setup.—Consider a quantum system formed from $N \gg 1$ degrees of freedom. The Hamiltonian, H, is chaotic. It conserves a number $\ll N$ of charges Q_a that do not all commute: $[H, Q_a] = 0$, but $[Q_a, Q_{a'}] \neq 0$ for some $a' \neq a$. H lacks other symmetries.

We illustrate with N qubits whose global spin components $S_{a=x,y,z}$ are conserved. We omit $\hbar/2$ factors. H, $\vec{S}^2$, and S_z share an eigenbasis $\{|\alpha, m\rangle\}$:

$$H|\alpha, m\rangle = E_\alpha|\alpha, m\rangle, \tag{3}$$

$$\vec{S}^2|\alpha, m\rangle = s_\alpha(s_\alpha + 1)|\alpha, m\rangle, \quad \text{and} \tag{4}$$

$$S_z|\alpha, m\rangle = m|\alpha, m\rangle, \quad \text{wherein} \tag{5}$$

$$m = -s_\alpha, -s_\alpha + 1, \ldots, s_\alpha\,. \tag{6}$$

Ladder operators $S_\pm = S_x \pm iS_y$ raise and lower S_z. Since H is chaotic and lacks other symmetries, $E_\alpha = E_{\alpha'} \Leftrightarrow \alpha = \alpha'$.

Our results govern the thermodynamic limit ($N \to \infty$) and the physically realizable mesoscale (large but finite N). The correspondence principle (as systems grow large, they grow classical) suggests that effects of noncommutation (a nonclassical phenomenon) vanish in the thermodynamic limit. Therefore, the mesoscale is particularly interesting here.

The normalized initial state decomposes as

$$|\psi(0)\rangle = \sum_{\alpha,m} C_{\alpha,m}|\alpha, m\rangle\,, \quad \text{wherein} \quad C_{\alpha,m} \in \mathbb{C}\,. \tag{7}$$

Operators $\mathscr{O}$ have time-t expectation values $\langle\mathscr{O}\rangle_t := \langle\psi(t)|\mathscr{O}|\psi(t)\rangle$. We drop the subscript from time constants:

$$\langle H\rangle =: E\,, \quad \text{and} \tag{8}$$

$$\langle S_z\rangle =: M\,. \tag{9}$$

Aligning the z-axis with $\langle\vec{S}\rangle$, we set $\langle S_x\rangle, \langle S_y\rangle = 0$. The state has an extensive energy, $E = O(N)$, and is far from maximally spin-polarized: $N - M = O(N)$.[3]

The system begins in an *approximate microcanonical subspace*, which generalizes a microcanonical subspace for noncommuting charges [22, 35, 48]. Measuring any charge Q_a likely yields an outcome near $\langle Q_a\rangle$; the charges' variances are bounded as

$$\mathrm{var}(H) \leq O(N), \tag{10}$$

$$\mathrm{var}(S_z) \leq O(N), \quad \text{and} \tag{11}$$

$$\mathrm{var}(S_{x,y}) \leq O(N)\,. \tag{12}$$

[2] We focus on continuous symmetries to bridge ETH studies with the emerging subfield of the quantum thermodynamics of noncommuting charges (Hermitian operators that generate continuous symmetries).

[3] ETH-type statements tend to hold when the thermodynamic entropy is extensive. S_{th} is generally nonextensive when additive conserved quantities (e.g., E or the S_a's) lie near their extremes, which we exclude.

Conditions (10)–(12) govern typical many-body states prepared today, including all short-range-correlated states [35, 48].[4]

Having introduced the initial state, we profile observables expected to obey the non-Abelian ETH. Blessed with a symmetry, we focus on symmetry-adapted operators, without sacrificing generality: *Spherical tensor operators* consist of components $T_q^{(k)}$ that can linearly combine to form any observable $\mathscr{O}$ [51]. For example, consider an atom absorbing a photon (of spin $k = 1$), which imparts q=1 quantum of z-type angular momentum. $T_{q=1}^{(k=1)}$ represents the photon's effect. Generally, the index $q = -k, -k+1, \ldots, k$; and the operator obeys the commutation relations $[S_z, T_q^{(k)}] = q\, T_q^{(k)}$ and $[S_\pm, T_q^{(k)}] = \sqrt{(k \mp q)(k \pm q + 1)}\, T_{q\pm1}^{(k)}$. Examples include spin operators that act nontrivially on a site j and as the identity elsewhere: The spin component $s_{j,z}$ is a $T_0^{(1)}$, and the ladder operators $s_{j,\pm}$ are $T_{\pm1}^{(1)}$.

We focus on *K-local observables* $\mathscr{O}$, formed from tensoring together $\leq K$ single-qubit factors $\mathscr{O}_j$, potentially summing such products, and normalizing such that $\mathscr{O}$'s operator norm is $\leq O(K)$.[5] Every K-local observable equals a linear combination of spherical-tensor components $T_q^{(\leq K)}$. We focus on $K = O(1)$, so $k = O(1)$, so $q = O(1)$. Such observables are expected to satisfy an ETH-type postulate [1, 52].

Non-Abelian ETH.— Consider representing an observable $T_q^{(k)}$ as a matrix relative to the energy eigenbasis. The matrix elements obey the Wigner–Eckart theorem [51],

$$\langle \alpha, m | T_q^{(k)} | \alpha', m' \rangle = \langle s_\alpha, m | s'_\alpha, m'; k, q \rangle \langle \alpha || T^{(k)} || \alpha' \rangle . \tag{13}$$

The Clebsch–Gordan coefficient $\langle s_\alpha, m | s'_\alpha, m'; k, q \rangle$ is nonzero only if $m' + q = m$ and $s_\alpha = |s'_\alpha - k|, |s'_\alpha - k| + 1, \ldots, s'_\alpha + k$; in the photon-and-atom example, the particles' initial and final quantum numbers obey the rules of angular-momentum addition. The reduced matrix element $\langle \alpha || T^{(k)} || \alpha' \rangle$ can obey the *non-Abelian ETH*, we posit now.

Define the average energy $\mathcal{E} \coloneqq \frac{1}{2}(E_\alpha + E_{\alpha'})$, energy difference $\omega \coloneqq E_\alpha - E_{\alpha'}$, and matrix R as in the introduction. Analogously, define the average spin quantum number $\mathcal{S} \coloneqq \frac{1}{2}(s_\alpha + s'_\alpha)$ and the difference $\nu \coloneqq s_\alpha - s'_\alpha$. Denote by $S_{\rm th}(\mathcal{E}, \mathcal{S})$ the thermodynamic entropy at energy $\mathcal{E}$ and spin quantum number $\mathcal{S}$. The observable $T_q^{(k)}$ and Hamiltonian H obey the non-Abelian ETH if, for smooth, real functions $\mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})$ and $f_\nu^{(k)}(\mathcal{E}, \mathcal{S}, \omega)$,[6]

$$\begin{aligned} \langle \alpha || T^{(k)} || \alpha' \rangle = {} & \mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})\, \delta_{\alpha,\alpha'} \\ & + e^{-S_{\rm th}(\mathcal{E},\mathcal{S})/2}\, f_\nu^{(k)}(\mathcal{E}, \mathcal{S}, \omega) R_{\alpha\alpha'} \, . \end{aligned} \tag{14}$$

Unlike $\mathcal{E}$, $\mathcal{S}$ is nonextensive, so the $\mathcal{S}$ dependencies may be unexpected. Yet the Wigner–Eckart theorem (13) prevents $\langle \alpha || T_q^{(k)} || \alpha' \rangle$ from depending on m or m'. Hence the non-Abelian ETH can encode non-Abelian-charge conservation only through $\mathcal{S}$.

Thermal prediction.—Chaotic systems thermalize to the canonical state $\rho_{\rm can} \propto e^{-\beta H}$ if just energy is conserved; to the grand canonical state $\rho_{\rm GC} \propto e^{-\beta(H - \mu\mathcal{N})}$ if the energy and particle number $\mathcal{N}$ are conserved; etc. Which thermal state emerges depends on the charges. If they fail to commute, derivations of the thermal state's form break down [20, 22]. Certain derivations were generalized in quantum-information thermodynamics to accommodate noncommuting charges [21–23, 53, 54], leading to the *non-Abelian thermal state* (NATS),

$$\rho_{\rm NATS} \coloneqq e^{-\beta(H - \sum_a \mu_a Q_a)} / Z \, . \tag{15}$$

β and the effective chemical potentials μ_a are defined by (i) $\mathrm{Tr}(H \rho_{\rm NATS}) = E$ and (ii) the fixing of the charge expectation values $\mathrm{Tr}(Q_a\, \rho_{\rm NATS}) = \langle Q_a \rangle$ [35].[7] The partition function is $Z \coloneqq \mathrm{Tr}(e^{-\beta(H - \sum_a \mu_a Q_a)})$. States of the form (15) are often called *generalized Gibbs ensembles*, especially if H is integrable and the Q_a commute [55–57]. Signatures of $\rho_{\rm NATS}$ emerged dynamically in numerical simulations [35] and a trapped-ion experiment [48]; yet full thermalization to $\rho_{\rm NATS}$ has not been observed in a closed quantum many-body system. Furthermore, noncommuting charges were conjectured to alter thermalization [35]. Using the non-Abelian ETH, we identify wide classes of local observables and initial states that thermalize. However, we find opportunities for anomalously large deviations from thermal predictions in the mesoscale.

Our z-axis choice simplifies $\rho_{\rm NATS}$ to $e^{-\beta(H - \mu S_z)}/Z$. Although $\rho_{\rm NATS}$ shares its mathematical form with $\rho_{\rm GC}$, the physics differs significantly. Here, energy and three noncommuting charges are conserved globally and transported locally; during grand canonical thermalization, only energy and particles (commuting charges) are. In the grand canonical case, the global system begins in a microcanonical subspace. Here, no nontrivial microcanonical subspace (associated with $s_\alpha \neq 0$) exists, and

[4] Let d denote the spatial dimensionality. Equations (11) and (12) are satisfied if spin–spin correlations $\langle s_{j,a} s_{j',a} \rangle - \langle s_{j,a} \rangle \langle s_{j',a} \rangle$ decay more quickly than $|j - j'|^{-d}$ as the spatial separation $|j - j'| \to \infty$. If this latter condition governs energy-density correlations, Eq. (10) holds.

[5] Example K-local observables include densities $\frac{1}{N} \sum_{j=1}^{N-1} \mathscr{O}_j \mathscr{O}_{j+1} \ldots \mathscr{O}_{j+K-1}$.

[6] Specifically, $\mathcal{T}^{(k)}$ and $f_\nu^{(k)}$ are smooth functions of $\mathcal{E}/N$ and $\mathcal{S}/N$ in the thermodynamic limit.

[7] In a non-Abelian twist on chemical potential, the μ_a transform as an adjoint representation of SU(2). If rotating bodies replace the spins, the μ_a reduce to angular velocities normalized by β.

the global system begins in an approximate microcanonical subspace. These differences in setup permit differences in thermalization.

$T_q^{(k)}$ has a thermal expectation value $\langle T_q^{(k)} \rangle_{\rm th} := {\rm Tr}(T_q^{(k)} \rho_{\rm NATS})$ whose trace we calculate using the $|\alpha, m\rangle$ basis. We apply the Wigner–Eckart theorem (13), then the non-Abelian ETH (14). The Clebsch–Gordan coefficient vanishes if $q \neq 0$, so[8]

$$\left\langle T_q^{(k)} \right\rangle_{\rm th} = \frac{\delta_{q,0}}{Z} \sum_{\alpha,m} e^{-\beta(E_\alpha - \mu m)} \, \langle s_\alpha, m | s_\alpha, m; k, 0 \rangle \times \mathcal{T}^{(k)}(E_\alpha, s_\alpha) \, . \tag{16}$$

Note that $\rho_{\rm NATS}$ satisfies the variance conditions (10)–(12) by Laplace's method and, independently, by the state's being short-range-correlated [58].4

Time-averaged expectation value.—After $|\psi(0)\rangle$ [Eq. (7)] evolves for a time t, the observable $T_q^{(k)}$ has an expectation value

$$\left\langle T_q^{(k)} \right\rangle_t = \sum_{\alpha,\alpha',m,m'} C^*_{\alpha,m} C_{\alpha',m'} \, e^{i(E_\alpha - E_{\alpha'})t} \times \langle \alpha, m | T_q^{(k)} | \alpha', m' \rangle . \tag{17}$$

We invoke the Wigner–Eckart theorem (13), invoke the non-Abelian ETH (14), and average $\langle T_q^{(k)} \rangle_t$ over an infinite time ($\lim_{t\to\infty} \frac{1}{t} \int_0^t dt'$). For all $\alpha' \neq \alpha$, the exponential in (17) dephases, so the "off-diagonal" terms vanish:

$$\overline{\langle T_q^{(k)} \rangle_t} = \sum_{\alpha,m} C^*_{\alpha,m+q} C_{\alpha,m} \langle s_\alpha, m+q | s_\alpha, m; k, q \rangle \times \mathcal{T}^{(k)}(E_\alpha, s_\alpha) \, . \tag{18}$$

Comparison.—We prove two results: (i) If $M = O(N)$, the time average (18) equals the thermal average (16), plus $O(N^{-1})$ corrections. Such corrections emerge also in the absence of noncommuting charges [50]. (ii) If $M = 0$, the time average may deviate from the thermal average by anomalously large, $O(N^{-1/2})$ corrections. These corrections appear sourced by different physics: quantum uncertainty in noncommuting charges, rather than thermodynamic ensembles' distinguishability at finite N [50].[9]

Consider an extensive $M = O(N)$ and $s_{j,z}$-like observables $T_{q=0}^{(k)}$. The thermal average (16) and time average (18) both assume the form

$$\sum_{\alpha,m} p_{\alpha,m} \langle s_\alpha, m | s_\alpha, m; k, 0 \rangle \, \mathcal{T}^{(k)}(E_\alpha, s_\alpha) =: \left\langle T_0^{(k)} \right\rangle_p . \tag{19}$$

The probabilities $p_{\alpha,m}$ are unit-normalized and have bounded moments: For all nonvanishing nonnegative-integer triples $(A, B, C) \in (\mathbb{Z}_{\geq 0})^3 \setminus (0,0,0)$,

$$\left\langle (E_\alpha - E)^A \, (m - M)^B \, (s_\alpha - M)^C \right\rangle_p \leq O\left(N^{A+B+C-1} \right) . \tag{20}$$

For a variable X, we have defined $\langle X \rangle_p$ as the average over $\{p_{\alpha,m}\}$. The distribution equals $\{e^{-\beta(E_\alpha - \mu m)}/Z\}$ in the thermal average and, in the time average, the diagonal ensemble $\{|C_{\alpha,m}|^2\}$. Both distributions satisfy the moment condition (20) by the scalings $E, M = O(N)$ and the variance conditions (10)–(12) (App. A).

In Eq. (19), we can Taylor-approximate the averaged quantity about ($E_\alpha = E$, $m = M$, $s_\alpha = M$), thanks to the moment condition (20) and to the Clebsch–Gordan coefficient's and $\mathcal{T}^{(k)}$'s smoothness (App. B). In the Taylor expansion, a general term is an $n^{\rm th}$-order derivative times an $n^{\rm th}$-order moment, for some $n \geq 0$. The derivative is $\leq O(N^{-n})$, by the functions' smooth dependence on E_α/N, m/N, and s_α/N. By the moment condition (20), the moment is $\leq O(N^{n-1})$. Hence Eq. (19) reduces to

$$\langle T_0^{(k)} \rangle_p = \mathcal{T}^{(k)}(E, M) + O(N^{-1}) \tag{21}$$

(App. B). Therefore, the thermal and time averages are equal to within $O(N^{-1})$ corrections.

Now, consider ladder-operator-like observables $T_{q\neq 0}^{(k)}$. The thermal average (16) vanishes, due to the Kronecker delta. We upper-bound the time average (18) in App. C, using the Cauchy-Schwarz inequality; a large-s_α approximation of Clebsch–Gordan coefficients (App. D); and, for $q = \pm 1$, the assumption $\langle S_{x,y} \rangle_0 = 0$. The time average is $\leq O(N^{-1})$, equaling the vanishing thermal average to within $O(N^{-1})$ corrections.

We now identify two cases where the corrections can be anomalously large. In both, $M = 0$, so $\mu = 0$. First, consider any rotationally invariant observable $T_0^{(0)}$. The Clebsch–Gordan coefficient $\langle s_\alpha, m | s_\alpha, m; 0, 0 \rangle = 1$. Hence the thermal average (16) and time average (18) reduce to

$$\left\langle T_0^{(0)} \right\rangle_{\rm th} = \frac{1}{Z} \sum_{\alpha,m} e^{-\beta E_\alpha} \, \mathcal{T}^{(0)}(E_\alpha, s_\alpha) \quad \text{and} \tag{22}$$

$$\overline{\left\langle T_0^{(0)} \right\rangle_t} = \sum_{\alpha,m} |C_{\alpha,m}|^2 \, \mathcal{T}^{(0)}(E_\alpha, s_\alpha) \, . \tag{23}$$

Appendix E details the argument sketched here: We Taylor-expand the smooth function $\mathcal{T}^{(0)}(E_\alpha, s_\alpha)$ about ($E_\alpha = E$, $s_\alpha = 0$). In the expansion, a general term is a moment $\langle (E_\alpha - E)^A \, (s_\alpha)^C \rangle_p$ times a derivative of $\mathcal{T}^{(0)}$.

[8] We consistently drop corrections exponentially small in N.

[9] Anomalous thermalization may occur also at intermediate scalings $M = O(N^\gamma)$, wherein $0 < \gamma < 1$. (Indeed, our $M = 0$ constructions can be modified to show anomalous thermalization $\forall \, \gamma \leq 1/2$.) However, this range encompasses cases too numerous for this paper. It suffices to show that (i) K-local observables thermalize under most circumstances but (ii) anomalous thermalization is possible under a physically reasonable assumption.

The leading ($A = C = 0$) term averages to $\mathcal{T}^{(0)}(E, 0)$. Terms with $A \geq 1$ or $C \geq 2$ average to $\leq O(N^{-1})$, as before. The remaining ($A = 0, C = 1$) term is linear in s_α/N. We crucially assume that this linear term is present—that the derivative $\partial\mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S})/\partial\mathcal{S}|_{\mathcal{S}=0}$ is nonzero. We argue, in App. F, that this is physically reasonable for suitable Hamiltonians and observables. In the thermal average (22), $\langle s_\alpha \rangle_{\text{th}} = O(N^{1/2})$, so the s_α/N term evaluates to $O(N^{-1/2})$. In contrast, in the time average (23), we can engineer $C_{\alpha,m}$ to be large only when $s_\alpha = O(1)$. The s_α/N term will time-average to $O(N^{-1})$. Hence the time average deviates from the thermal average by

$$\left\langle T_0^{(0)} \right\rangle_{\text{th}} - \overline{\langle T_0^{(0)} \rangle_t} = O(N^{-1/2}) > O(N^{-1}) \,. \tag{24}$$

Appendix G details a second scenario for anomalous thermalization. We construct a $|\psi(0)\rangle$ with four nonzero coefficients $C_{\alpha,m}$ [Eq. (7)], which have relative phases ± 1. The anomalously thermalizing observable is a $T_q^{(k)}$ whose $k > 0$ is even and whose $q = 1$. The argument relies on (i) a symmetry of Clebsch–Gordan coefficients and (ii) the smooth function $\mathcal{T}^{(k)}(E_\alpha, s_\alpha)$'s having a term of $O(s_\alpha/N)$, as above. The time average $\overline{\langle T_1^{(k)} \rangle_t}$ deviates from the thermal average $\langle T_1^{(k)} \rangle_{\text{th}}$ by an $O(N^{-1/2})$ correction, as in (24).

Anomalous $O(N^{-1/2})$ scaling characterizes also a result in [22]. Reference [22] generalized a conventional derivation of the thermal state's form to accommodate noncommuting charges. The global system was assumed to be in an approximate microcanonical state (a normalized projector onto an approximate microcanonical subspace). The environment was traced out. The reduced state's distance to ρ_{NATS} was upper-bounded with (const.)$N^{-1/2}$ + (const.), echoing Eq. (24). Hence our results, based on the ETH and dynamics, may reflect the kinematic results in [22].

Outlook.—We have extended the eigenstate thermalization hypothesis, a cornerstone of many-body physics, to the more fully quantum scenario in which conserved charges fail to commute with each other. Noncommutation can prevent the charges from sharing an eigenspace (a microcanonical subspace) and invalidates the usual assumption that the Hamiltonian lacks degeneracies. We overcame these challenges by proposing a non-Abelian ETH and focusing on an approximate microcanonical subspace. Applying these tools to SU(2), we compared the long-time average of an observable's expectation value with the thermal expectation value. The averages agree in many cases, e.g., whenever $M = O(N)$. Yet the averages can disagree by anomalously large $O(N^{-1/2})$ corrections otherwise, under a physically reasonable assumption.

This work opens several avenues for future research. First, our analytical results call for testing with numerics and quantum simulators, e.g., trapped ions, ultracold atoms, and superconducting qudits [47, 48]. One would aim to verify the non-Abelian ETH (14); identify operators whose smooth functions $\mathcal{T}^{(k)}$ satisfy our assumptions, enabling anomalous thermalization; and observe deviations (24) from the thermal prediction. Second, we expect our arguments to generalize from SU(2). Third, the smooth function $f_\nu^{(k)}(\mathcal{E}, \mathcal{S}, \omega)$ [Eq. (14)] should reveal how non-Abelian symmetries influence thermalization dynamics. Fourth, the thermodynamics of noncommuting charges [19–45, 47, 48] merits bridging to many-body scars, which also have emergent non-Abelian symmetries and violate the ETH [59, 60]. This work extends the ETH to the more fully quantum regime of noncommuting charges, linking many-body physics to quantum-information thermodynamics.

ACKNOWLEDGMENTS

N.Y.H. thanks Wen Wei Ho, Noah Lupu-Gladstein, Mikhail Lukin, and Bihui Zhu, and C.M. thanks Tibor Rakovszky and Curt von Keyserlingk, for helpful conversations. C.M. and N.Y.H. acknowledge the hospitality of the Kavli Institute for Theoretical Physics, where this work started. N.Y.H. acknowledges also the Perimeter Institute's hospitality. This research was supported in part by the Gordon and Betty Moore Foundation's EPiQS Initiative through GBMF8686 (C.M.), by the National Science Foundation under QLCI grant OMA-2120757 (N.Y.H.), by an NSF Graduate Research Fellowship under Grant No. 2139319 (F.I.), and by the National Science Foundation under Grant No. NSF PHY-1748958. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities.

Appendix A THE INITIAL STATE SATISFIES THE MOMENT CONDITION.

The main text casts the diagonal ensemble $\{|C_{\alpha,m}|^2\}$ as satisfying the moment condition (20),

$$\left\langle (E_\alpha - E)^A \, (m - M)^B \, (s_\alpha - M)^C \right\rangle_p \leq O\left(N^{A+B+C-1}\right) \quad \forall (A, B, C) \in (\mathbb{Z}_{\geq 0})^3 \setminus (0, 0, 0) \, . \tag{A1}$$

We prove the claim here. The proof relies on (i) the finite dimensionality of the local subsystems' Hilbert spaces; (ii) the scalings

$$E = O(N) \quad \text{and} \quad M = O(N) \, ; \tag{A2}$$

and (iii) the variance conditions (10)–(12), repeated here for convenience:

$$\text{var}(H) = \langle H^2 \rangle - E^2 \leq O(N), \tag{A3}$$

$$\text{var}(S_z) = \langle S_z^2 \rangle - M^2 \leq O(N), \quad \text{and} \tag{A4}$$

$$\text{var}(S_{x,y}) = \langle S_{x,y}^2 \rangle \leq O(N) \, . \tag{A5}$$

The variance conditions are satisfied, for example, if $|\psi(0)\rangle$ is short-range-correlated.[4]

The proof proceeds in three steps. First, we derive analogs for $\vec{S}^2$ of the scaling and variance conditions (A2) and (A3)–(A5). Second, we upper-bound fairly general correlators' magnitudes. Third, we combine steps 1–2.

Step 1: Unlike S_z, $\vec{S}^2$ is nonextensive. Therefore, $\vec{S}^2$'s expectation value and variance are not necessarily bounded as S_z's are. However, we can prove similar bounds. We define the diagonal average of any variable $X_{\alpha,m}$ as

$$\langle X_{\alpha,m} \rangle_{\text{diag}} \coloneqq \sum_{\alpha,m} |C_{\alpha,m}|^2 X_{\alpha,m}. \tag{A6}$$

We will prove the bounds

$$\langle s_\alpha \rangle_{\text{diag}} \leq M + O(1) \quad \text{and} \quad \left\langle (s_\alpha - M)^2 \right\rangle_{\text{diag}} \leq O(N) \, . \tag{A7}$$

First, we sum the variance conditions (A4) and (A5):

$$\langle \vec{S}^2 \rangle - M^2 \leq O(N). \tag{A8}$$

We evaluate the left-hand side (LHS) on $|\psi(0)\rangle$ [Eq. (7)]:

$$\sum_{\alpha,m} |C_{\alpha,m}|^2 s_\alpha (s_\alpha + 1) - M^2 \leq O(N). \tag{A9}$$

The inequality is equivalent, by algebra and the normalization of $\{|C_{\alpha,m}|^2\}$, to

$$M + (2M + 1) \sum_{\alpha,m} |C_{\alpha,m}|^2 (s_\alpha - M) + \sum_{\alpha,m} |C_{\alpha,m}|^2 (s_\alpha - M)^2 \leq O(N) \, . \tag{A10}$$

Since $M = \sum_{\alpha,m} |C_{\alpha,m}|^2 \, m$, we can replace the second term's $(s_\alpha - M)$ with $(s_\alpha - m)$. Recall that $m \leq s_\alpha$. Every factor on the inequality's LHS is therefore nonnegative, so every term is, so every term must be $\leq O(N)$. Since $2M + 1 = O(N)$, the second term implies that $\sum_{\alpha,m} |C_{\alpha,m}|^2 (s_\alpha - M) \leq O(1)$. By the definition of $\langle . \rangle_{\text{diag}}$, we recover the first inequality in (A7). The second inequality in (A7) follows similarly from Ineq. (A10)'s third term.

Step 2: We now upper-bound fairly general correlators' magnitudes. Let $x_1, x_2, \ldots, x_n$ denote real-valued functions of α and m. (For notational brevity, we suppress the functions' dependencies on α and m.) Let the functions' magnitudes obey the upper bound $|x_j| \leq X \in \mathbb{R} \ \forall j, \alpha, m$. We analyze correlator magnitudes of the form

$$\left| \left\langle x_1^{A_1} x_2^{A_2} \ldots x_n^{A_n} \right\rangle_{\text{diag}} \right| . \tag{A11}$$

Without loss of generality, the powers are ordered from greatest to least: $A_1 \geq A_2 \geq \ldots \geq A_n \geq 0$. At-least-two-point correlators interest us, so $A \coloneqq \sum_{j=1}^n A_j \geq 2$. Therefore, either $A_1 \geq 2$ or $A_1 = A_2 = 1$. In the first case, we show,

the correlator magnitude (A11) is upper-bounded by X^{A-2} times $\langle x_1^2 \rangle$; in the second case, the correlator magnitude is upper-bounded by X^{A-2} times $\frac{1}{2}\langle x_1^2 + x_2^2 \rangle$. We parcel the factors so for reasons clarified in step 3.

First, suppose that $A_1 \geq 2$. To upper-bound (A11), we invoke the average's definition, then the triangle inequality:

$$\left| \left\langle x_1^{A_1} x_2^{A_2} \dots x_n^{A_n} \right\rangle_{\text{diag}} \right| \leq \sum_{\alpha,m} |C_{\alpha,m}|^2 \, |x_1|^{A_1} |x_2|^{A_2} \cdots |x_n|^{A_n} . \tag{A12}$$

We separate out a factor of $|x_1|^2$. Then, we bound the rest using the assumption $|x_j| \leq X$ and the definition $A \coloneqq \sum_{j=1}^n A_j$:

$$\left| \left\langle x_1^{A_1} x_2^{A_2} \dots x_n^{A_n} \right\rangle_{\text{diag}} \right| \leq \sum_{\alpha,m} |C_{\alpha,m}|^2 \, |x_1|^2 \cdot \underbrace{|x_1|^{A_1-2} |x_2|^{A_2} \cdots |x_n|^{A_n}}_{\leq X^{A-2}} \leq \langle x_1^2 \rangle_{\text{diag}} \, X^{A-2} . \tag{A13}$$

The final inequality follows from the reality of x_1.

Now, suppose that $A_1 = A_2 = 1$. To bound the correlator magnitude (A11), we again invoke the average's definition, then the triangle inequality. This time, we separate $x_1^{A_1} x_2^{A_2} = x_1 x_2$ from the other variables:

$$\left| \left\langle x_1^{A_1} x_2^{A_2} \dots x_n^{A_n} \right\rangle_{\text{diag}} \right| \leq \sum_{\alpha,m} |C_{\alpha,m}|^2 \, |x_1 x_2| \cdot \underbrace{|x_3|^{A_3} |x_4|^{A_4} \cdots |x_n|^{A_n}}_{\leq X^{A-2}} . \tag{A14}$$

Since x_1 and x_2 are real, $x_1^2 + x_2^2 - 2|x_1 x_2| = (|x_1| - |x_2|)^2 \geq 0$. Rearranging yields $|x_1 x_2| \leq \frac{1}{2}(x_1^2 + x_2^2)$. Combining this inequality with Ineq. (A14), we obtain

$$\left| \left\langle x_1^{A_1} x_2^{A_2} \dots x_n^{A_n} \right\rangle_{\text{diag}} \right| \leq \frac{1}{2} \left\langle x_1^2 + x_2^2 \right\rangle_{\text{diag}} X^{A-2} . \tag{A15}$$

Step 3: We now synthesize steps 1 and 2. Let (x_1, x_2, x_3) equal $(E_\alpha - E,\, m - M,\, s_\alpha - M)$ or some permutation thereof. Since local subsystems have finite-dimensional Hilbert spaces, each variable is upper-bounded by some $O(N)$ number X. By the variance conditions, the functions $\langle x_j^2 \rangle_{\text{diag}}$ and $\frac{1}{2}\langle x_j^2 + x_k^2 \rangle_{\text{diag}}$ are $O(N)$ for all $j, k = 1, 2, 3$. Therefore, substituting into Eq. (A13) yields the moment condition (A1), as does substituting into Eq. (A15).

We can now explain why, during step 2, we sought bounds that contained $\langle x_1^2 \rangle_{\text{diag}}$ or $\langle x_1^2 + x_2^2 \rangle_{\text{diag}}$. These averages are only $O(N)$. If we had treated x_1^2 or $x_1 x_2$ like the other variables, each would have contributed an $O(N^2)$ factor to the corresponding bound. We would not have recovered the all-important -1 in the moment condition's exponent [Eq. (A1)].

Appendix B CALCULATION OF TIME AVERAGE AND THERMAL AVERAGE WHEN $M = O(N)$ AND $q = 0$

We aim to evaluate the LHS of Eq. (19),

$$\sum_{\alpha,m} p_{\alpha,m} \langle s_\alpha, m | s_\alpha, m; k, 0 \rangle \, \mathcal{T}^{(k)}(E_\alpha, s_\alpha) =: \left\langle T_0^{(k)} \right\rangle_p . \tag{B1}$$

The moment condition (A1) implies that $p_{\alpha,m}$ peaks near $(E_\alpha = E,\, m = M,\, s_\alpha = M)$. Furthermore, the Clebsch–Gordan coefficient and $\mathcal{T}^{(k)}$ are smooth. Hence we can Taylor-approximate each function about $(E_\alpha = E,\, m = M,\, s_\alpha = M)$.

Taylor expansion of Clebsch–Gordan coefficient: In App. D, we approximate the Clebsch–Gordan coefficients at $s_\alpha \gg 1$ and $s_\alpha - m \ll s_\alpha$. The latter condition does govern the dominant contributions to Eq. (B1): The moment condition (A1), together with $M = O(N)$, implies that $\langle (s_\alpha - m)^n / s_\alpha^n \rangle_p \leq O(N^{-1})$, for $n \geq 1$. To prove this inequality, we Taylor-expand $(s_\alpha - m)^n / s_\alpha^n$ about $m = M$ and $s_\alpha = M$. A general term in the expansion is of $O\big([m - M]^B [s_\alpha - M]^C / M^{B+C}\big)$, wherein $B, C \in \mathbb{Z}_{\geq 0}$ and $B + C \geq n$. Each such term averages to $\leq O(N^{-1})$. Hence

$$\left\langle \left[\frac{s_\alpha - m}{s_\alpha} \right]^n \right\rangle_p \leq O\left(N^{-1}\right) , \quad \text{for} \quad n \geq 1. \tag{B2}$$

Having justified the use of the asymptotic expansion in App. D, we now use the expansion. Substituting $q = 0$ into Eq. (D13) yields

$$\langle s_\alpha, m | s_\alpha, m; k, 0\rangle = 1 + O\left(\frac{s_\alpha - m}{s_\alpha}\right) + \dots \tag{B3}$$

The $\dots$ consists of terms that contain additional powers of $(s_\alpha - m)/s_\alpha$. Taylor-expanding the Clebsch–Gordan coefficient about $m = s_\alpha = M$, which is $O(N)$, yields

$$\langle s_\alpha, m | s_\alpha, m; k, 0\rangle = 1 + O\left(\frac{m - M}{N}\right) + O\left(\frac{s_\alpha - M}{N}\right) + \dots \tag{B4}$$

A general term in the expansion is of $O\big([m - M]^B [s_\alpha - M]^C / N^{B+C}\big)$, wherein $B, C \in \mathbb{Z}_{\geq 0}$.

Taylor expansion of $\mathcal{T}^{(k)}$: By assumption, $\mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})$ is a smooth function of $\mathcal{E}/N$ and $\mathcal{S}/N$. Therefore,

$$\frac{\partial^A}{\partial \mathcal{E}^A} \frac{\partial^C}{\partial \mathcal{S}^C} \mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})\bigg|_{\mathcal{E}=E, \mathcal{S}=M} = O\left(\frac{1}{N^{A+C}}\right). \tag{B5}$$

Hence the Taylor expansion of $\mathcal{T}^{(k)}(E_\alpha, s_\alpha)$ about $(E_\alpha = E,\ s_\alpha = M)$ has the form

$$\mathcal{T}^{(k)}(E_\alpha, s_\alpha) = \mathcal{T}^{(k)}(E, M) + O\left(\frac{E_\alpha - E}{N}\right) + O\left(\frac{s_\alpha - M}{N}\right) + \dots \tag{B6}$$

A general term in the expansion is of $O\left([E_\alpha - E]^A [s_\alpha - M]^C / N^{A+C}\right)$, wherein $A, C \in \mathbb{Z}_{\geq 0}$.

Combining the two Taylor expansions: We substitute the Taylor series (B4) and (B6) into the average (B1), then multiply out. As discussed in the main text's setup section, $\mathcal{T}^{(k)}(E, M) = O(1)$. Therefore,

$$\left\langle T_0^{(k)} \right\rangle_p = \sum_{\alpha, m} p_{\alpha,m} \left[\mathcal{T}^{(k)}(E, M) + O\left(\frac{E_\alpha - E}{N}\right) + O\left(\frac{m - M}{N}\right) + O\left(\frac{s_\alpha - M}{N}\right) + \dots \right]. \tag{B7}$$

A general term has the form $O\big([E_\alpha - E]^A [m - M]^B [s_\alpha - M]^C / N^{A+B+C}\big)$. In the leading term, $\mathcal{T}^{(k)}(E, M)$ can be factored out of the sum, which then equals one, by the normalization of $\{p_{\alpha,m}\}$. The general remaining term averages, by the moment condition (A1), to $\leq O(N^{A+B+C-1}/N^{A+B+C}) = O(N^{-1})$. Hence

$$\langle T_0^{(k)} \rangle_p = \mathcal{T}^{(k)}(E, M) + O(N^{-1}), \tag{B8}$$

as quoted in Eq. (21).

Appendix C CALCULATION OF TIME AVERAGE WHEN $M = O(N)$ AND $q \neq 0$

Let us upper-bound the time-averaged expectation value (18), assuming that $M = O(N)$ and $q \neq 0$:

$$\overline{\left\langle T_{q\neq 0}^{(k)} \right\rangle}_t = \sum_{\alpha, m} C^*_{\alpha, m+q} C_{\alpha, m} \langle s_\alpha, m + q | s_\alpha, m; k, q\rangle\, \mathcal{T}^{(k)}(E_\alpha, s_\alpha). \tag{C1}$$

When q vanished (App. B), we simplified the average using properties of $|C_{\alpha,m}|^2$. We can achieve some of that simplification here, using the Cauchy-Schwarz inequality: Define the vectors $\vec{u}$ and $\vec{v}$ in terms of the components $u^*_{\alpha,m} = C^*_{\alpha,m+q}\sqrt{\langle s_\alpha, m + q | s_\alpha, m; k, q\rangle \mathcal{T}^{(k)}(E_\alpha, s_\alpha)}$ and $v_{\alpha,m} = C_{\alpha,m}\sqrt{\langle s_\alpha, m + q | s_\alpha, m; k, q\rangle \mathcal{T}^{(k)}(E_\alpha, s_\alpha)}$. Any branch-cut convention can be applied to the square root. Define the inner product $\vec{u} \cdot \vec{v} \coloneqq \sum_{\alpha,m} u^*_{\alpha,m} v_{\alpha,m}$. The Cauchy-Schwarz inequality states that $|\vec{u} \cdot \vec{v}| \leq \sqrt{\vec{u} \cdot \vec{u}}\sqrt{\vec{v} \cdot \vec{v}} \leq \max\{\vec{u} \cdot \vec{u},\ \vec{v} \cdot \vec{v}\}$, so

$$\left|\overline{\left\langle T_{q\neq 0}^{(k)} \right\rangle}_t\right| \leq \max\left\{ \sum_{\alpha,m} |C_{\alpha,m}|^2 \left|\langle s_\alpha, m | s_\alpha, m - q; k, q\rangle \mathcal{T}^{(k)}(E_\alpha, s_\alpha)\right|, \right.$$
$$\left. \sum_{\alpha,m} |C_{\alpha,m}|^2 \left|\langle s_\alpha, m + q | s_\alpha, m; k, q\rangle \mathcal{T}^{(k)}(E_\alpha, s_\alpha)\right| \right\}. \tag{C2}$$

We have redefined $m + q \mapsto m$ in the first sum.

The sums are dominated by terms in which

$$(E_\alpha \sim E,\, m \sim M,\, s_\alpha \sim M) \quad \text{and so, by the moment condition (A1)}, \quad s_\alpha - m \ll s_\alpha \,, \tag{C3}$$

as when $q = 0$ (App. B). We approximate the Clebsch–Gordan coefficients under these conditions in App. D. The result [Eqs. (D13) and (D14)] is

$$\langle s_\alpha, m+q | s_\alpha, m; k, q\rangle = O\left(\left[\frac{s_\alpha - m}{s_\alpha}\right]^{|q|/2}\right) + \ldots \tag{C4}$$

The $\ldots$ consists of terms that contain additional powers of $(s_\alpha - m)/s_\alpha$ or $1/s_\alpha$. The same asymptotic expansion characterizes the $\langle s_\alpha, m | s_\alpha, m-q; k, q\rangle$ in Eq. (C2).

We substitute the expansion (C4) into Ineq. (C2). Since $\mathcal{T}^{(k)}(E_\alpha, s_\alpha) \leq O(1)\,,$

$$\left|\overline{\left\langle T^{(k)}_{q\neq 0}\right\rangle_t}\right| \leq \sum_{\alpha,m} |C_{\alpha,m}|^2 \left\{ O\left(\left[\frac{s_\alpha - m}{s_\alpha}\right]^{|q|/2}\right) + \ldots \right\} . \tag{C5}$$

The $\ldots$ consists of terms that contain additional powers of $(s_\alpha - m)/s_\alpha$ or $1/s_\alpha$. We evaluate the bound for $|q| \geq 2\,,$ $q = 1\,,$ and $q = -1$ sequentially.

Bounding the time-averaged expectation value when $|q| \geq 2$: The moment condition (A1), together with $M = O(N)$, implies that $\langle (s_\alpha - m)^n / s_\alpha^n \rangle_{\rm diag} \leq O(N^{-1})$ for $n \geq 1$ [Eq. (B2)]. Therefore, since the $C_{\alpha,m}$'s are normalized to one, Ineq. (C5) implies that

$$\left|\overline{\left\langle T^{(k)}_{q}\right\rangle_t}\right| \leq O(N^{-1}) \quad \text{if} \quad |q| \geq 2\,. \tag{C6}$$

The time average equals the thermal average (zero), to within $O(N^{-1})$ corrections.

Bounding the time-averaged expectation value when $q = +1$: We return to the bound (C5). Since $q = 1\,,$ the leading term averages to $\langle [(s_\alpha - m)/s_\alpha]^{1/2} \rangle_{\rm diag} = O(N^{-1/2})$ in Eq. (C2). Therefore, we cannot immediately conclude that the time average $\leq O(N^{-1})\,.$

To demonstrate the time average's smallness, we return to Eq. (C1). We expect the same terms to dominate as when $|q| \geq 2$ [Eq. (C3)]. Accordingly, Eq. (D13) approximates the Clebsch–Gordan coefficient:

$$\langle s_\alpha, m+1 | s_\alpha, m; k, 1\rangle = -\sqrt{\frac{k(k+1)}{2}}\,\sqrt{\frac{s_\alpha - m}{s_\alpha}}\left[1 + O\left(\frac{s_\alpha - m}{s_\alpha}\right) + \ldots\right] . \tag{C7}$$

Substituting into Eq. (C1) yields

$$\overline{\left\langle T^{(k)}_{q=1}\right\rangle_t} = -\sqrt{\frac{k(k+1)}{2}} \sum_{\alpha,m} C^*_{\alpha,m+1} C_{\alpha,m} \sqrt{\frac{s_\alpha - m}{s_\alpha}}\,\left[1 + O\left(\frac{s_\alpha - m}{s_\alpha}\right) + \ldots\right] \mathcal{T}^{(k)}(E_\alpha, s_\alpha)\,. \tag{C8}$$

We can Taylor-expand everything except the $\sqrt{s_\alpha - m}$ about $(E_\alpha = E,\, m = M,\, s_\alpha = M)\,.$ Since $\mathcal{T}^{(k)}(E_\alpha, s_\alpha) = \mathcal{T}^{(k)}(E, M) + O\big(\frac{E_\alpha - E}{N}\big) + O\big(\frac{s_\alpha - M}{N}\big) + \ldots$ [Eq. (B6)],

$$\begin{aligned}\overline{\left\langle T^{(k)}_{q=1}\right\rangle_t} = -\sqrt{\frac{k(k+1)}{2}} &\sum_{\alpha,m} C^*_{\alpha,m+1} C_{\alpha,m} \sqrt{\frac{s_\alpha - m}{M}} \\ &\times \left[\mathcal{T}^{(k)}(E, M) + O\left(\frac{E_\alpha - E}{N}\right) + O\left(\frac{m - M}{N}\right) + O\left(\frac{s_\alpha - M}{N}\right) + \ldots\right] .\end{aligned} \tag{C9}$$

The leading term, involving $\mathcal{T}^{(k)}(E, M)$, is the problematic one [for proving that the time average is $\leq O(N^{-1})$]: The leading term looks to be of $O(N^{-1/2})$. However, this term is actually proportional to $\langle S_+ \rangle \equiv \langle S_x \rangle + i\,\langle S_y \rangle$, which vanishes by assumption. To prove the proportionality, we substitute the initial-state expansion $|\psi(0)\rangle = \sum_{\alpha,m} C_{\alpha,m} |\alpha, m\rangle$ and the raising-operator equation $S_+ |\alpha, m\rangle = \sqrt{(s_\alpha - m)(s_\alpha + m + 1)}\,|\alpha, m+1\rangle$ into the expectation value $\langle S_+ \rangle = \sum_{\alpha,m} C^*_{\alpha,m+1} C_{\alpha,m} \sqrt{(s_\alpha - m)(s_\alpha + m + 1)}\,.$ Taylor-approximating $\sqrt{s_\alpha + m + 1} = \sqrt{2M+1}\,\big[1 + O\big(\frac{s_\alpha - M}{N}\big) + O\big(\frac{m - M}{N}\big) + \ldots\big]$ yields

$$\langle S_+ \rangle = \sqrt{2M+1} \sum_{\alpha,m} C^*_{\alpha,m+1} C_{\alpha,m} \sqrt{s_\alpha - m}\,\left[1 + O\left(\frac{s_\alpha - M}{N}\right) + O\left(\frac{m - M}{N}\right) + \ldots\right] = 0\,. \tag{C10}$$

The leading-order term is proportional to the leading-order term in Eq. (C9). Therefore, at leading order, $\overline{\left\langle T^{(k)}_{q=1} \right\rangle_t} = 0$. The higher-order terms in Eq. (C9) can be shown to evaluate to $\leq O(N^{-1})$; one repeats the Cauchy-Schwarz argument used for $|q| \geq 2$. Therefore,

$$\left| \overline{\left\langle T^{(k)}_{q=1} \right\rangle_t} \right| \leq O(N^{-1}) \,. \tag{C11}$$

Bounding the time-averaged expectation value when $q = -1$: The proof is almost the same as for $q = 1$. Instead of $\langle S_+ \rangle = 0$, we use $\langle S_- \rangle \equiv \langle S_x \rangle - i \langle S_y \rangle = 0$:

$$\left| \overline{\left\langle T^{(k)}_{q=-1} \right\rangle_t} \right| \leq O(N^{-1}) \,. \tag{C12}$$

Thus, for all $q \neq 0$, the time average equals the thermal average (zero), to within $O(N^{-1})$ corrections.

Appendix D APPROXIMATION OF CLEBSCH–GORDAN COEFFICIENTS

Here, we approximate the Clebsch–Gordan coefficients $\langle s_\alpha, m+q | s_\alpha, m; k, q \rangle$ when $s_\alpha \gg 1$ and $s_\alpha - m \ll s_\alpha$. As throughout this paper, $k, q = O(1)$.

The general expression for Clebsch–Gordan coefficients is [61, Eq. (2.41)]

$$\begin{aligned} \langle s, m | s', m'; k, q \rangle &= \delta_{m,\, m'+q} \\ &\times \sqrt{\frac{(2s+1)\,(s+s'-k)!\,(s-s'+k)!\,(s'+k-s)!\,(s+m)!\,(s-m)!\,(s'-m')!\,(s'+m')!\,(k-q)!\,(k+q)!}{(s+s'+k+1)!}} \\ &\times \sum_\ell \frac{(-1)^\ell}{\ell!\,(s'+k-s-\ell)!\,(s'-m'-\ell)!\,(k+q-\ell)!\,(s-k+m'+\ell)!\,(s-s'-q+\ell)!} \,. \end{aligned} \tag{D1}$$

The final line's sum runs over all integer ℓ values for which every factorial's argument is nonnegative. This expression holds for $m \geq 0$ and $s' > k$. We set $s' = s$, as in the time-averaged expectation value (18). We also set $m = m' + q$ and drop the primes:

$$\begin{aligned} \langle s, m+q | s, m; k, q \rangle = &\sqrt{\frac{(2s+1)\,(2s-k)!\,(k!)^2\,(s+m+q)!\,(s-m-q)!\,(s-m)!\,(s+m)!\,(k-q)!\,(k+q)!}{(2s+k+1)!}} \\ &\times \sum_\ell \frac{(-1)^\ell}{\ell!\,(k-\ell)!\,(s-m-\ell)!\,(k+q-\ell)!\,(s-k+m+\ell)!\,(\ell-q)!} \,. \end{aligned} \tag{D2}$$

This expression holds for $m + q \geq 0$ and $s > k$. Both conditions are satisfied in the regime of interest, wherein $s \gg 1$ and $s - m \ll s$, while $k, q = O(1)$. The sum runs over the integers ℓ for which each factorial's argument is nonnegative.

Together, the factorials imply three upper bounds and three lower bounds on ℓ:

$$\ell \leq k \,, \tag{D3}$$
$$\ell \leq s - m \,, \tag{D4}$$
$$\ell \leq k + q \,, \tag{D5}$$
$$\ell \geq k - s - m \,, \tag{D6}$$
$$\ell \geq q \,, \quad \text{and} \tag{D7}$$
$$\ell \geq 0 \,. \tag{D8}$$

We initially assume that $q \geq 0$. Consequently, Ineq. (D3) subsumes Ineq. (D5), and Ineq. (D7) subsumes Ineq. (D8). As $s \gg 1$ and $s - m \ll s$, Ineq. (D6) encodes a trivially negative lower bound. The constraints on ℓ reduce to

$$\ell \in \big\{ q, q+1, \ldots, \min\{k, s-m\} \big\} \,. \tag{D9}$$

Let us extract the asymptotics of the Clebsch–Gordan coefficient in the limit as $s \to \infty$, assuming $(s-m)/s \to 0$. For convenience, we change variables from m to $\Delta \coloneqq s - m$. Grouping the s-dependent factors together yields

$$\langle s, m+q | s, m; k, q \rangle = \sum_{\ell} \frac{(-1)^{\ell}\, k!\, \sqrt{(\Delta-q)!\, \Delta!\, (k-q)!\, (k+q)!}}{\ell!\, (k-\ell)!\, (\Delta-\ell)!\, (k+q-\ell)!\, (\ell-q)!} \times \sqrt{\frac{(2s+1)(2s-k)!(2s-\Delta+q)!\, (2s-\Delta)!}{(2s+k+1)!}}\, \frac{1}{(2s-\Delta-k+\ell)!}\,. \tag{D10}$$

The s-dependent factorials are all large in the limit of interest, so we approximate them using Stirling's formula:

$$x! = \exp\!\left(x \ln x - x + \frac{1}{2} \ln(2\pi x) + O(1/x) + \dots \right). \tag{D11}$$

We take the s-dependent expression's natural log; expand in powers of $1/s$; and exponentiate. The s-dependent factor is

$$(2s)^{q/2-\ell}\, \exp\Big(O(\Delta/s) + \cdots \Big) = (2s)^{q/2-\ell} \Big[1 + O(\Delta/s) + \cdots \Big]. \tag{D12}$$

Therefore, the least possible ℓ value dominates the $\sum_{\ell}$ in Eq. (D10). By (D9), that ℓ value is q.

Let us approximate the $\sum_{\ell}$ with the $\ell = q$ term, while replacing the s-dependent factor with (D12). We revert notation from Δ to $s - m$. The result is

$$\langle s, m+q | s, m; k, q \rangle = \frac{(-1)^q}{q! (2s)^{q/2}} \left(\frac{(s-m)!\, (k+q)!}{(s-m-q)!\, (k-q)!} \right)^{1/2} \left[1 + O\!\left(\frac{s-m}{s} \right) + \dots \right], \qquad q \geq 0\,. \tag{D13}$$

Now, suppose that $q < 0$. The bounds (D9) on ℓ become $\ell \in \big\{ 0, 1, \dots, \min\{k-q, s-m\} \big\}$. In Eq. (D10), $\ell = 0$ labels the sum's dominant term. The s-dependent factor approximates to $(2s)^{q/2}[1 + O\left(\frac{s-m}{s}\right) + \dots]$. Again, we substitute into and approximate Eq. (D10). The result is

$$\langle s, m+q | s, m; k, q \rangle = \frac{1}{|q|! (2s)^{|q|/2}} \left(\frac{(s-m+|q|)!\, (k+|q|)!}{(s-m)!\, (k-|q|)!} \right)^{1/2} \left[1 + O\!\left(\frac{s-m}{s} \right) + \dots \right], \qquad q < 0\,. \tag{D14}$$

Appendix E THERMAL AVERAGE OF ROTATIONALLY INVARIANT OBSERVABLES AT $M = 0$

The main text illustrated potential anomalous thermalization first with rotationally invariant operators $T_0^{(0)}$. Here, we evaluate in greater detail the thermal average (22),

$$\Big\langle T_0^{(0)} \Big\rangle_{\text{th}} = \frac{1}{Z} \sum_{\alpha, m} e^{-\beta E_\alpha}\, \mathcal{T}^{(0)}(E_\alpha, s_\alpha)\,. \tag{E1}$$

The summand is a smooth function of E_α/N and s_α/N when $N \gg 1$, by assumption. Therefore, we can replace the sum over states with an integral over energy and spin quantum numbers (treated as continuous variables), weighted by the density of states $e^{S_{\text{th}}}$:

$$\Big\langle T_0^{(0)} \Big\rangle_{\text{th}} \approx \frac{1}{Z'} \int_{E_{\min}}^{E_{\max}} d\mathcal{E} \int_0^{s_{\max}} d\mathcal{S}\; e^{S_{\text{th}}(\mathcal{E}, \mathcal{S}) - \beta \mathcal{E}}\, \mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S})\,. \tag{E2}$$

The normalization condition $\langle \mathbb{1} \rangle_{\text{th}} = 1$ fixes the effective partition function Z'.

We evaluate the integral using Laplace's method (the saddle-point approximation, but for real variables). For most fixed values of $\mathcal{E}/N$ and $\mathcal{S}/N$, the exponent $S_{\text{th}}(\mathcal{E}, \mathcal{S}) - \beta \mathcal{E} = O(N)$, so the integrand peaks steeply about this function's global maximum. The exponential's argument is stationary where the argument's first derivatives vanish:

$$\frac{\partial S_{\text{th}}}{\partial \mathcal{E}} - \beta = 0\,, \quad \text{and} \quad \frac{\partial S_{\text{th}}}{\partial \mathcal{S}} = 0. \tag{E3}$$

The first condition is the usual thermodynamic definition of the inverse temperature β. The thermodynamic entropy's concavity suggests that (E3) has a unique solution $(\mathcal{E}, \mathcal{S}) = (\mathcal{E}_*, \mathcal{S}_*)$, at which the exponent attains its global maximum. We Taylor-approximate the exponent about this maximum:

$$S_{\rm th}(\mathcal{E}, \mathcal{S}) - \beta\mathcal{E} \approx S_{\rm th}(\mathcal{E}_*, \mathcal{S}_*) - \beta\mathcal{E}_* + \frac{1}{2} \left(\mathcal{E} - \mathcal{E}_*, \, \mathcal{S} - \mathcal{S}_*\right)^T \left[\nabla_{\rm H}^2 S_{\rm th}\right]_* \left(\mathcal{E} - \mathcal{E}_*, \, \mathcal{S} - \mathcal{S}_*\right) . \tag{E4}$$

$\left[\nabla_{\rm H}^2 S_{\rm th}\right]_*$ denotes the Hessian matrix of $S_{\rm th}(\mathcal{E}, \mathcal{S})$, evaluated at $(\mathcal{E}_*, \mathcal{S}_*)$. Equation (E2) reduces to

$$\left\langle T_0^{(0)} \right\rangle_{\rm th} \approx \frac{1}{Z''} \int_{E_{\rm min}}^{E_{\rm max}} d\mathcal{E} \int_0^{s_{\rm max}} d\mathcal{S} \; e^{-\frac{1}{2} \left(\mathcal{E} - \mathcal{E}_*, \, \mathcal{S} - \mathcal{S}_*\right)^T \left[-\nabla_{\rm H}^2 S_{\rm th}\right]_* \left(\mathcal{E} - \mathcal{E}_*, \, \mathcal{S} - \mathcal{S}_*\right)} \mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S}) . \tag{E5}$$

We have absorbed into the effective partition function Z'' the leading term in the expansion (E4). Again, Z'' is fixed by the normalization condition $\langle \mathbb{1} \rangle_{\rm th} = 1$.

We can approximate, and assess the scalings of, several components of the integral (E5). First, according to conventional thermodynamics, $-\partial^2 S_{\rm th} / \partial \mathcal{E}^2 \propto 1/(\text{heat capacity})$ is positive and $O(N^{-1})$. Therefore, we expect $\left[-\nabla_{\rm H}^2 S_{\rm th}\right]_*$ to be a positive-definite matrix whose elements are $O(N^{-1})$. Therefore, the Gaussian factor in (E5) has a peak width of $O(N^{1/2})$. Second, by evaluating the LHS of $\langle H \rangle_{\rm th} \equiv E$, using Eq. (E5), we obtain $\mathcal{E}_* \approx E$. Third, by evaluating the LHS of $\langle \vec{S}^2 \rangle_{\rm th} = O(N)$,[10] also using Eq. (E5), we infer that $\mathcal{S}_* \in [0, \, O(N^{1/2})]$. Fourth, by the foregoing observations, the Gaussian peaks far from the integration limits $\mathcal{E} = E_{\rm min}, E_{\rm min}$ and $\mathcal{S} = s_{\rm max}$. Therefore, we can extend these limits to $\pm\infty$. Applying these conclusions to Eq. (E5) yields

$$\left\langle T_0^{(0)} \right\rangle_{\rm th} \approx \frac{1}{Z''} \int_{-\infty}^{\infty} d\mathcal{E} \int_0^{\infty} d\mathcal{S} \; e^{-\frac{1}{2} \left(\mathcal{E} - E, \, \mathcal{S} - \mathcal{S}_*\right)^T \left[-\nabla_{\rm H}^2 S_{\rm th}\right]_* \left(\mathcal{E} - E, \, \mathcal{S} - \mathcal{S}_*\right)} \mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S}) . \tag{E6}$$

Whereas the exponential peaks sharply, $\mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S})$ is smooth and varies slowly, by assumption. We therefore Taylor-expand $\mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S})$ about $\mathcal{E} = E$ and $\mathcal{S} = 0$:[11]

$$\mathcal{T}^{(0)}\left(\mathcal{E}, \mathcal{S}\right) \approx \mathcal{T}^{(0)}\left(E, 0\right) + O\left(\frac{\mathcal{E} - E}{N}\right) + O\left(\frac{\mathcal{S}}{N}\right) + \ldots \tag{E7}$$

We argue for the nonzero $O(\mathcal{S}/N)$ term's presence, for suitable Hamiltonians and suitable observables $\mathcal{T}^{(0)}$, in App. F.

We substitute the Taylor approximation (E7) into Eq. (E6), then perform standard multivariate Gaussian integration. The zeroth-order term in Eq. (E7) integrates to $\mathcal{T}^{(0)}(E, 0)$, by the integral's normalization. The $O([\mathcal{E} - E]/N)$ term integrates to zero, by the $\mathcal{E}$ integral's symmetry. The $O(\mathcal{S}/N)$ term does not integrate to zero similarly, because the $\mathcal{S}$ integral ends at $\mathcal{S} = 0$. However, the Gaussian has a width of $O(N^{1/2})$. Hence $\mathcal{S}$ averages to $O(N^{1/2})$, and the $O(\mathcal{S}/N)$ term in Eq. (E7) averages to $O(N^{-1/2})$:

$$\left\langle T_0^{(0)} \right\rangle_{\rm th} \approx \mathcal{T}^{(0)}\left(E, 0\right) + O\left(N^{-1/2}\right) . \tag{E8}$$

Appendix F TAYLOR EXPANSION OF $\mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})$ ABOUT $\mathcal{S} = 0$

The potential anomalous thermalization relies on two claims about the smooth function $\mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})$: Suppose that $k \geq 0$ is even and $\mathcal{E} = O(N)$. For some Hamiltonians H and observables $T_q^{(k)}$, $\mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S})$ can have the Taylor approximation about $\mathcal{S} = 0$

$$\mathcal{T}^{(k)}(\mathcal{E}, \mathcal{S}) = O(1) + O(\mathcal{S}/N) . \tag{F1}$$

Furthermore, if $k > 0$, then the $O(1)$ term vanishes. Here, we argue for the claims. If $k > 0$, a deductive argument supports the lack of an $O(1)$ term. No such arguments preclude (i) the $O(1)$ term when $k = 0$ or (ii) the $O(\mathcal{S}/N)$

[10] We evaluate $\langle \vec{S}^2 \rangle_{\rm th}$ by replacing the $\mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S})$ in Eq. (E5) with $\mathcal{S}(\mathcal{S} + 1)$. We can understand the $O(N)$ through the thermal state's sharing of scaling behaviors with the initial state: $\langle \vec{S}^2 \rangle_{\rm th} \sim \langle \vec{S}^2 \rangle_0 = \left\langle S_x^2 + S_y^2 + S_z^2 \right\rangle_0 = {\rm var}_0(S_x) + {\rm var}_0(S_y) + {\rm var}_0(S_z) + \langle S_z \rangle_0^2$. Each of the first three terms is $O(N)$, by the variance conditions (11)–(12). The final term vanishes because $M = 0$ in this appendix. Hence $\left\langle \vec{S}^2 \right\rangle_{\rm th} = O(N)$.

[11] Strictly speaking, one should Taylor-expand $\mathcal{T}^{(0)}(\mathcal{E}, \mathcal{S})$ about the maximum, $\mathcal{S} = \mathcal{S}_*$, rather than about $\mathcal{S} = 0$. However, $\mathcal{T}^{(0)}$ is sufficiently smooth, and $\mathcal{S}_* \in [0, O(N^{1/2})]$ is sufficiently close to 0, that the two expansions yield identical results.

term ever, to our knowledge. Hence there is no reason to believe that such terms are absent. Beyond this "everything not forbidden is compulsory" reasoning, we also argue for the $O(\mathcal{S}/N)$ term's plausibility.

Argument for the absence of any $O(1)$ term when $k > 0$: When $s_\alpha = 0$, the system lacks spin angular momentum and so any preferred direction. Therefore, all rotationally noninvariant operators' expectation values must vanish. $T_q^{(k)}$ is rotationally noninvariant for all q, because $k > 0$. To identify the implications for $\mathcal{T}^{(k)}(\mathcal{E}, 0)$, we evaluate $\langle \alpha, m | T_q^{(k)} | \alpha, m \rangle$ (which vanishes) on an $s_\alpha = 0$ eigenstate $|\alpha, m\rangle$. We invoke the Wigner–Eckart theorem (13) and the non-Abelian ETH (14). The associated Clebsch–Gordan coefficient equals one. Hence $\langle \alpha, m | T_q^{(k)} | \alpha, m \rangle = \mathcal{T}^{(k)}(E_\alpha, 0)$. The LHS vanishes, as argued above. Hence the $O(1)$ term $\mathcal{T}^{(k)}(\mathcal{E}, 0) = 0$.

Argument for the $O(\mathcal{S}/N)$ term in Eq. (F1)*:* Bound states underlie this argument. To provide intuition, we temporarily address a more familiar setting: a lattice in which only the global particle-number operator, $\mathcal{N}$, is conserved. Consider a global state $|\psi\rangle$ of uniformly distributed two-particle bound states. Wherever a particle appears, another particle appears beside it. Denote by $\mathcal{N}_j$ the site-j particle-number operator. For an arbitrary j, we estimate the correlator $\langle \psi | \mathcal{N}_j \mathcal{N}_{j+1} | \psi \rangle$.

The correlator equals the joint probability p(particle at site j, particle at site $j+1$). Semiclassically, this joint probability equals p(particle at site $j+1$|particle at site j) $\times$ p(particle at site j). The latter probability equals $O(\langle \psi | \mathcal{N} | \psi \rangle / N)$, by the state's uniformity. The conditional probability is $O(1)$, because the particles are bound. Hence the joint probability p(particle at site j, particle at site $j+1$) $= O(\langle \psi | \mathcal{N} | \psi \rangle / N)$.

We reason about noncommuting charges by analogy with the preceding argument. First, suppose that $k \geq 2$. Let the Hamiltonian have a finite-energy-density eigenstate $|\alpha, m{=}s_\alpha\rangle$ that contains bound clusters of k z-type charges. For example, ferromagnetic couplings can cause neighboring spins to point in the same direction. If that direction is $\hat{z}$, the state contains bound z-charges. If H has some degree of uniformity, so can the energy eigenstate. s_α is essentially the amount of charge in the global system.

A local observable of interest has the form

$$T_0^{(k)} \sim s_{j_1,z}\, s_{j_2,z} \ldots s_{j_k,z} \,. \tag{F2}$$

Recall that $s_{j,z}$ denotes qubit j's z-type spin operator. This $T_0^{(k)}$ has an expectation value, in a joint eigenstate $|\alpha, m{=}s_\alpha\rangle$, that is essentially a k-point correlator:

$$\langle \alpha, s_\alpha | T_0^{(k)} | \alpha, s_\alpha \rangle \sim \langle \alpha, s_\alpha | s_{j_1,z}\, s_{j_2,z} \ldots s_{j_k,z} | \alpha, s_\alpha \rangle \,. \tag{F3}$$

Similarly to in the particle-number example, this correlator is essentially the joint probability

$$\begin{aligned} &p(\text{site } j_1 \text{ contains a quantum of } z\text{-type charge, site } j_2 \text{ contains a quantum of } z\text{-charge}, \ldots, \\ &\quad \text{site } j_k \text{ contains a quantum of } z\text{-charge}) \,. \end{aligned} \tag{F4}$$

Semiclassically, this joint probability equals

$$\begin{aligned} &p(\text{site } j_2 \text{ contains a quantum of } z\text{-charge}, \ldots, \text{site } j_k \text{ contains a quantum of } z\text{-charge} \\ &\quad | \text{site } j_1 \text{ contains a quantum of } z\text{-charge}) \times p(\text{site } j_1 \text{ contains a quantum of } z\text{-charge}) \,. \end{aligned} \tag{F5}$$

The final probability is $O(s_\alpha/N)$, by the state's uniformity. The conditional probability is $O(1)$, if j_1 lies close to the other j's, because the charges form bound clusters. Hence the joint probability is $O(s_\alpha/N)$. So, semiclassically, $\langle \alpha, s_\alpha | T_0^{(k)} | \alpha, s_\alpha \rangle = O(s_\alpha/N)$. The LHS is essentially $\mathcal{T}^{(k)}(E_\alpha, s_\alpha)$, by the Wigner–Eckart theorem (13) and the non-Abelian ETH (14). Hence $\mathcal{T}^{(k)}(E_\alpha, s_\alpha) = O(s_\alpha/N)$.

If $k = 0$, a similar argument concerns a local observable $T_0^{(0)} = \vec{s}_j \cdot \vec{s}_{j'}$. Start with a "resonating valence bond" (RVB) state, formed from superpositions of local dimer coverings (local pairings of the qubits into singlets). This state has zero total spin ($s_\alpha = 0$). The expectation value $\langle T_0^{(0)} \rangle = \langle -\vec{s}_j \cdot \vec{s}_{j'} \rangle$ quantifies the probability that qubits j and j' form a singlet. This probability is $O(1)$ in the RVB state, if j lies close to j'. Imagine that a few of the singlets are broken—a small fraction ρ. Furthermore, the broken singlets are distributed uniformly throughout the system. The resulting state has nonzero but small total spin: $s_\alpha/N \sim \rho$. The expectation value $\langle T_0^{(0)} \rangle = \langle \vec{s}_j \cdot \vec{s}_{j'} \rangle = O(1) + O(s_\alpha/N)$. The second term equals the probability that the j-and-j' singlet is broken. Hence, by the same logic as for $k \geq 2$, $\mathcal{T}^{(0)}(E_\alpha, s_\alpha) = O(1) + O(s_\alpha/N)$.

Appendix G SECOND OPPORTUNITY FOR ANOMALOUS THERMALIZATION

This appendix details a second opportunity for anomalous thermalization when $M = 0$. We construct a time-independent state $|\psi(0)\rangle$ that has the properties stipulated in our setup. Then, we focus on observables $T^{(k)}_{q=1}$, for an even $k \geq 2$. The time average $\overline{\langle T^{(k)}_1 \rangle_t}$, we show, differs from the thermal average $\langle T^{(k)}_1 \rangle_{\rm th}$ at the anomalously large $O(N^{-1/2})$.

Consider an initial state $|\psi(0)\rangle$ in a Hamiltonian eigenspace labeled by $\alpha = A$. Let the eigenenergy $E_A = O(N)$ and spin quantum number $s_A = O(N^{1/2})$. For some to-be-specified magnetic spin quantum number $\bar{m} \in [-s_A, s_A]$,

$$|\psi(0)\rangle \coloneqq \frac{1}{2}(|A, \bar{m}\rangle + |A, \bar{m}+1\rangle + |A, -\bar{m}\rangle - |A, -\bar{m}-1\rangle)\,. \tag{G1}$$

One can check directly that $M = \langle S_z \rangle = 0$ and $\mathrm{var}(S_z) = \frac{1}{2}[\bar{m}^2 + (\bar{m}+1)^2] = O(N)$. That $\langle S_{x,y} \rangle = 0$ follows from (i) the decompositions of S_x and S_y in terms of $S_\pm$ and (ii) the ladder operators' actions on an S_z eigenstate, $S_\pm|A, m\rangle = \sqrt{s_A(s_A+1) - m(m\pm 1)}\,|A, m\pm 1\rangle$. The same ingredients imply that $\langle S^2_{x,y} \rangle = O(s^2_A) + O(\bar{m}^2) = O(N)$; hence $\mathrm{var}(S_{x,y}) = O(N)$. The energy variance, $\mathrm{var}(H)$, vanishes by construction. Hence $|\psi(0)\rangle$ has the properties stipulated in the main text's setup section, including the variance conditions (10)–(12).

Having prescribed an initial state, we shift focus to an observable $T^{(k)}_q$. Let $q = 1$. To calculate the time-averaged expectation value, we substitute the $C_{\alpha,m}$'s from Eq. (G1) into Eq. (18):

$$\overline{\left\langle T^{(k)}_1 \right\rangle_t} = \frac{1}{4}\Big(\langle s_A, \bar{m}+1 | s_A, \bar{m}; k, 1\rangle - \langle s_A, -\bar{m} | s_A, -\bar{m}-1; k, 1\rangle\Big)\, \mathcal{T}^{(k)}(E_A, s_A)\,. \tag{G2}$$

The Clebsch–Gordan coefficients obey the symmetry relation

$$\langle s_\alpha, m+1 | s_\alpha, m; k, 1\rangle = (-1)^{k+1} \langle s_\alpha, -m | s_\alpha - m - 1; k, 1\rangle \tag{G3}$$

[61, Eq. (2.42)]. Consequently, if k is odd, the time average (G2) vanishes. Since $S_\pm$ are $T^{(1)}_{\pm 1}$ operators, we have corroborated our earlier conclusion that $\langle S_{x,y} \rangle = 0$.

Suppose that k is greater than 0 (recall that $k = 0$ in the first opportunity for anomalous thermalization). and is even. The symmetry (G3) reduces the time average (G2) to

$$\overline{\left\langle T^{(k)}_1 \right\rangle_t} = \frac{1}{2}\langle s_A, \bar{m}+1 | s_A, \bar{m}; k, 1\rangle\, \mathcal{T}^{(k)}(E_A, s_A)\,. \tag{G4}$$

In App. F, we Taylor-approximated the smooth function $\mathcal{T}^{(k)}(E_\alpha, s_\alpha)$ about $s_\alpha = 0$. The Taylor approximation, we argued, can have the form $\mathcal{T}^{(k)}(E_A, s_A) = O(s_\alpha/N) \neq 0$. We can choose $\bar{m}$ to make Eq. (G4)'s Clebsch–Gordan coefficient be $O(1)$. In fact, the Clebsch–Gordan coefficient is $O(1)$ for most choices of $\bar{m} \in [-s_A, s_A]$ that are not too close to $\pm s_A$. Therefore, substituting into Eq. (G4) yields

$$\overline{\left\langle T^{(k)}_1 \right\rangle_t} = O\left(N^{-1/2}\right)\,. \tag{G5}$$

The thermal average $\langle T^{(k)}_1 \rangle_{\rm th}$ vanishes, because $q \neq 0$, by Eq. (16). Hence the time average differs from the thermal average at $O(N^{-1/2})$.

REFERENCES

[1] L. D'Alessio, Y. Kafri, A. Polkovnikov, and M. Rigol, Advances in Physics **65**, 239 (2016).
[2] J. M. Deutsch, Phys. Rev. A **43**, 2046 (1991).
[3] M. Srednicki, Phys. Rev. E **50**, 888 (1994).
[4] M. Rigol, V. Dunjko, and M. Olshanii, Nature **452**, 854 (2008).
[5] L. Foini and J. Kurchan, Physical Review E **99**, 042139 (2019).
[6] S. Pappalardi, L. Foini, and J. Kurchan, arXiv preprint arXiv:2204.11679 (2022).
[7] J. M. Deutsch, Reports on Progress in Physics **81**, 082001 (2018).
[8] A. Chandran, M. D. Schulz, and F. J. Burnell, Phys. Rev. B **94**, 235122 (2016).
[9] D. Jansen, J. Stolpp, L. Vidmar, and F. Heidrich-Meisner, Phys. Rev. B **99**, 155130 (2019).
[10] T.-C. Lu and T. Grover, Phys. Rev. E **99**, 032111 (2019).

[11] C. Murthy and M. Srednicki, Phys. Rev. E **100**, 022131 (2019).
[12] T. Langen, R. Geiger, and J. Schmiedmayer, Annual Review of Condensed Matter Physics **6**, 201 (2015).
[13] N. Lashkari, A. Dymarsky, and H. Liu, Journal of Statistical Mechanics: Theory and Experiment **2018**, 033101 (2018).
[14] N. Bao and N. Cheng, Journal of High Energy Physics **2019**, 152 (2019).
[15] C. Murthy and M. Srednicki, Phys. Rev. Lett. **123**, 230606 (2019).
[16] N. Mueller, T. V. Zache, and R. Ott, arXiv e-prints , arXiv:2107.11416 (2021), 2107.11416.
[17] M. Srednicki, Journal of Physics A: Mathematical and General **32**, 1163 (1999).
[18] L. D. Landau and E. M. Lifshitz, *Statistical Physics: Part 1* (Butterworth-Heinemann, 1980).
[19] M. Lostaglio, The resource theory of quantum thermodynamics, Master's thesis, Imperial College London, 2014.
[20] N. Yunger Halpern, Journal of Physics A: Mathematical and Theoretical **51**, 094001 (2018).
[21] Y. Guryanova, S. Popescu, A. J. Short, R. Silva, and P. Skrzypczyk, Nature Communications **7**, 12049 (2016).
[22] N. Yunger Halpern, P. Faist, J. Oppenheim, and A. Winter, Nat Commun **7**, 12051 (2016).
[23] M. Lostaglio, D. Jennings, and T. Rudolph, New Journal of Physics **19**, 043008 (2017).
[24] C. Sparaciari, L. Del Rio, C. M. Scandolo, P. Faist, and J. Oppenheim, Quantum **4**, 259 (2020).
[25] Z. B. Khanian, arXiv preprint arXiv:2012.14143 (2020).
[26] Z. B. Khanian, M. N. Bera, A. Riera, M. Lewenstein, and A. Winter, arXiv preprint arXiv:2011.08020 (2020).
[27] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, Nature Communications **9**, 5352 (2018).
[28] G. Manzano, J. M. Parrondo, and G. T. Landi, arXiv preprint arXiv:2011.04560 (2020).
[29] S. Popescu, A. B. Sainz, A. J. Short, and A. Winter, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **376**, 20180111 (2018).
[30] S. Popescu, A. B. Sainz, A. J. Short, and A. Winter, Physical Review Letters **125**, 090601 (2020).
[31] K. Ito and M. Hayashi, Phys. Rev. E **97**, 012129 (2018).
[32] M. N. Bera, A. Riera, M. Lewenstein, Z. B. Khanian, and A. Winter, Quantum **3**, 121 (2019).
[33] J. Mur-Petit, A. Relaño, R. A. Molina, and D. Jaksch, Nature Communications **9**, 2006 (2018).
[34] G. Manzano, Physical Review E **98**, 042123 (2018).
[35] N. Yunger Halpern, M. E. Beverland, and A. Kalev, Phys. Rev. E **101**, 042117 (2020).
[36] G. Manzano *et al.*, Physical Review Research **2**, 043302 (2020).
[37] K. Fukai, Y. Nozawa, K. Kawahara, and T. N. Ikeda, Physical Review Research **2**, 033403 (2020).
[38] J. Mur-Petit, A. Relaño, R. A. Molina, and D. Jaksch, arXiv preprint arXiv:1910.11000 (2019).
[39] M. Scandi and M. Perarnau-Llobet, Quantum **3**, 197 (2019).
[40] P. Boes, H. Wilming, J. Eisert, and R. Gallego, Nature communications **9**, 1 (2018).
[41] Y. Mitsuhashi, K. Kaneko, and T. Sagawa, arXiv e-prints , arXiv:2103.06060 (2021), 2103.06060.
[42] T. Croucher, J. Wright, A. R. R. Carvalho, S. M. Barnett, and J. A. Vaccaro, Information erasure, in *Thermodynamics in the Quantum Regime: Fundamental Aspects and New Directions*, edited by F. Binder, L. A. Correa, C. Gogolin, J. Anders, and G. Adesso, pp. 713–730, Springer International Publishing, Cham, 2018.
[43] J. A. Vaccaro and S. M. Barnett, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **467**, 1770 (2011).
[44] J. S. S. T. Wright, T. Gould, A. R. R. Carvalho, S. Bedkihal, and J. A. Vaccaro, Phys. Rev. A **97**, 052104 (2018).
[45] T. Croucher and J. A. Vaccaro, arXiv e-prints , arXiv:2111.10930 (2021), 2111.10930.
[46] I. Marvian, H. Liu, and A. Hulse, arXiv:2105.12877 (2021).
[47] N. Yunger Halpern and S. Majidy, npj Quantum Information **8**, 10 (2022).
[48] F. Kranzl *et al.*, arXiv e-prints , arXiv:2202.04652 (2022), 2202.04652.
[49] I. Marvian, H. Liu, and A. Hulse, arXiv e-prints , arXiv:2202.01963 (2022), 2202.01963.
[50] M. Srednicki, Journal of Physics A: Mathematical and General **29**, L75 (1996).
[51] R. Shankar, *Principles of Quantum Mechanics*, 2nd ed. (Spring, 2008).
[52] J. R. Garrison and T. Grover, Phys. Rev. X **8**, 021026 (2018).
[53] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[54] R. Balian and N. L. Balazs, Annals of Physics **179**, 97 (1987).
[55] M. Rigol, Phys. Rev. Lett. **103**, 100403 (2009).
[56] M. Rigol, V. Dunjko, V. Yurovsky, and M. Olshanii, Phys. Rev. Lett. **98**, 050405 (2007).
[57] L. Vidmar and M. Rigol, Journal of Statistical Mechanics: Theory and Experiment **2016**, 064007 (2016).
[58] M. Kliesch and A. Riera, *Properties of Thermal Quantum States: Locality of Temperature, Decay of Correlations, and More* (Springer International Publishing, Cham, 2018), pp. 481–502.
[59] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papic, Nature Physics **14**, 745 (2018).
[60] S. Choi *et al.*, Phys. Rev. Lett. **122**, 220603 (2019).
[61] A. Bohm and M. Loewe, *Quantum mechanics: Foundations and Applications*, 2nd ed. (Springer-Verlag New York, 1986), p. 172.